

Please type a plus sign (+) inside this box → ☐

PTO/SB/05 (12/97) (modified)  
Approved for use through 09/30/00. OMB 0651-0032  
Patent and Trademark Office: U.S. DEPARTMENT OF COMMERCE

**UTILITY  
PATENT APPLICATION  
TRANSMITTAL**

(Only for new nonprovisional applications under 37 CFR 1.53(b))

Attorney Docket No. 348022001200

Total Pages 57

First Named Inventor or Application Identifier

De Chao Yu and Daniel R. Henderson

Express Mail Label No. EL 034 669 514 US

**CERTIFICATE OF MAILING BY "EXPRESS MAIL"**

Express Mail Label No.: EL 034 669 514 US

Date of Deposit: September 9, 1999

I hereby certify that this paper or fee is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 C.F.R. § 1.10 on the date indicated above and is addressed to: Assistant Commissioner for Patents, Washington, D.C. 20231.

Shirley Laboy

**APPLICATION ELEMENTS**

See MPEP chapter 600 concerning utility patent application contents.

ADDRESS TO:

Assistant Commissioner for Patents  
Box Patent Application  
Washington, DC 20231

1. ☒ Fee Transmittal Form  
(Submit an original, and a duplicate for fee processing)
2. ☒ Specification [Total Pages 53]  
(preferred arrangement set forth below)
  - Descriptive title of the Invention
  - Cross References to Related Applications
  - Statement Regarding Fed sponsored R & D
  - Reference to Microfiche Appendix
  - Background of the Invention
  - Brief Summary of the Invention
  - Brief Description of the Drawings (if filed)
  - Detailed Description
  - Claim(s)
  - Abstract of the Disclosure
3. ☒ Drawing(s) (35 USC 113) [Total Sheets 27]
4. ☐ Oath or Declaration [Total Pages ]
  - a. ☐ Newly executed (original or copy)
  - b. ☐ Copy from a prior application (37 CFR 1.63(d)  
(for continuation/divisional with Box 17 completed)  
[Note Box 5 below]
  - i. ☐ DELETION OF INVENTOR(S)  
Signed statement attached deleting inventor(s) named in  
the prior application, see 37 CFR 1.63(d)(2) and 1.33(b)
5. ☐ Incorporation By Reference (useable if Box 4b is checked)  
The entire disclosure of the prior application, from which a copy of the  
oath or declaration is supplied under Box 4b, is considered as being  
part of the disclosure of the accompanying application and is hereby  
incorporated by reference therein.

6. ☐ Microfiche Computer Program (Appendix)
7. Nucleotide and/or Amino Acid Sequence Submission  
(if applicable, all necessary)
  - a. ☐ Computer Readable Copy
  - b. ☐ Paper Copy (identical to computer copy)
  - c. ☐ Statement verifying identity of above copies

**ACCOMPANYING APPLICATION PARTS**

8. ☐ Assignment Papers (cover sheet & document(s))
9. ☐ 37 CFR 3.73(b) Statement ☐ Power of Attorney  
(when there is an assignee)
10. ☐ English Translation Document (if applicable)
11. ☐ Information Disclosure Statement (IDS)/PTO-1449 ☐ Copies of IDS Citations
12. ☐ Preliminary Amendment
13. ☒ Return Receipt Postcard (MPEP 503)  
(Should be specifically itemized)
14. ☐ Small Entity ☐ Statement filed in prior application,  
Statement(s) Status still proper and desired
15. ☐ Certified Copy of Priority Document(s)  
(if foreign priority is claimed)
16. ☐

17. If a CONTINUING APPLICATION, check appropriate box and supply the requisite information:

☐ Continuation ☐ Divisional ☐ Continuation-in-part (CIP) of prior application No:

**18. CORRESPONDENCE ADDRESS**

Catherine M. Polizzi  
Registration No. 40,130

Morrison & Foerster LLP  
755 Page Mill Road  
Palo Alto, California 94304-1018  
Telephone: (650) 813-5651  
Facsimile: (650) 494-0792

- ☒ If a paper is untimely filed in the above-referenced application by applicant or his/her representative, the Assistant Commissioner is hereby petitioned under 37 C.F.R. § 1.136(a) for the minimum extension of time required to make said paper timely. In the event a petition for extension of time is made under the provisions of this paragraph, the Assistant Commissioner is hereby requested to charge any fee required under 37 C.F.R. § 1.17(a)-(d) to **Deposit Account No. 03-1952**. However, the Assistant Commissioner is **NOT** authorized to charge the cost of the issue fee to the Deposit Account.

The filing fee has been calculated as follows:

| FOR   | NUMBER FILED | NUMBER EXTRA | RATE       | CALCULATIONS |
|---|--------------|--------------|------------|--------------|
| TOTAL CLAIMS  | 28 - 20 =    | 8            | x \$18.00  | \$224.00     |
| INDEPENDENT CLAIMS  | 2 - 3 =      | 0            | x \$78.00  | \$0.00       |
| MULTIPLE DEPENDENT CLAIM(S) (if applicable)   |              |              | + \$260.00 | \$0.00       |
|   |              |              | BASIC FEE  | \$760.00     |
| TOTAL OF ABOVE CALCULATIONS =   |              |              |            | \$984.00     |
| Reduction by 1/2 for filing by small entity (Note 37 C.F.R. §§ 1.9, 1.27, 1.28).<br>If applicable, verified statement must be attached. |              |              |            | \$492.00     |
|   |              |              |            |              |
|   |              |              | TOTAL =    | \$492.00     |

Applicant(s) hereby petitions for any required relief including extensions of time and authorizes the Assistant Commissioner to charge the cost of such petitions and/or other fees or to credit any overpayment to **Deposit Account No. 03-1952** referencing docket no. 348022001200. A duplicate copy of this transmittal is enclosed, for that purpose.

Dated: September 9, 1999

Respectfully submitted,

By: Catherine M. Polizzi

Catherine M. Polizzi  
Registration No. 40,130

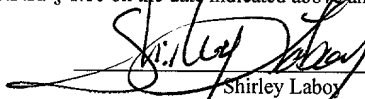
Morrison & Foerster LLP  
755 Page Mill Road  
Palo Alto, California 94304-1018  
Telephone: (650) 813-5651  
Facsimile: (650) 494-0792

**CERTIFICATE OF MAILING BY "EXPRESS MAIL"**

Express Mail Label No.: EL 034 669 514 US

Date of Deposit: September 9, 1999

I hereby certify that this paper or fee is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 C.F.R. § 1.10 on the date indicated above and is addressed to: Assistant Commissioner for Patents, Washington, D.C. 20231.

  
Shirley Laboy**ADENOVIRUS VECTORS CONTAINING CELL STATUS-SPECIFIC  
RESPONSE ELEMENTS AND METHODS OF USE THEREOF****CROSS-REFERENCE TO RELATED APPLICATIONS**

This application claims the priority benefit of U.S. Provisional Patent Application No.60/099,791, filed September 10, 1998. The priority application is hereby incorporated herein by reference in its entirety.

**STATEMENT OF RIGHTS TO INVENTIONS MADE UNDER  
FEDERALLY SPONSORED RESEARCH**

(Not Applicable)

**TECHNICAL FIELD**

This invention relates to cell transfection using adenoviral vectors. More specifically, it relates to cell status-specific replication of adenovirus vectors in cells, regardless of tissue or cell type.

**BACKGROUND ART**

In spite of numerous advances in medical research, cancer remains the second leading cause of death in the United States. In the industrialized nations, roughly one in five persons will die of cancer. Traditional modes of clinical care, such as surgical resection, radiotherapy and chemotherapy, have a significant failure rate, especially

for solid tumors. Neoplasia resulting in benign tumors can usually be completely cured by removing the mass surgically. If a tumor becomes malignant, as manifested by invasion of surrounding tissue, it becomes much more difficult to eradicate. Once a malignant tumor metastasizes, it is much less likely to be eradicated.

5           Excluding basal cell carcinoma, there are over one million new cases of cancer per year in the United States alone, and cancer accounts for over one half million deaths per year in this country. In the world as a whole, the five most common cancers are those of lung, stomach, breast, colon/rectum, and uterine cervix, and the total number of new cases per year is over 6 million.

10           Lung cancer is one of the most refractory of solid tumors because inoperable cases are up to 60% and the 5-year survival is only 13%. In particular, adenocarcinomas, which comprise about one-half of the total lung cancer cases, are mostly chemo-radioresistant. Colorectal cancer is the third most common cancer and the second leading cause of cancer deaths. Pancreatic cancer is virtually always fatal.  
15           Thus, current treatment prospects for many patients with these carcinomas are unsatisfactory, and the prognosis is poor.

          Hepatocellular carcinoma (HCC or malignant hepatoma) is one of the most common cancers in the world, and is especially problematic in Asia. Treatment prospects for patients with hepatocellular carcinoma are dim. Even with  
20           improvements in therapy and availability of liver transplant, only a minority of patients are cured by removal of the tumor either by resection or transplantation. For the majority of patients, the current treatments remain unsatisfactory, and the prognosis is poor.

          Breast cancer is one of the most common cancers in the United States, with an  
25           annual incidence of about 182,000 new cases and nearly 50,000 deaths. In the industrial nations, approximately one in eight women can expect to develop breast cancer. The mortality rate for breast cancer has remained unchanged since 1930. It has increased an average of 0.2% per year, but decreased in women under 65 years of

age by an average of 0.3% per year. See e.g., Marchant (1994) Contemporary Management of Breast Disease II: Breast Cancer, in: *Obstetrics and Gynecology Clinics of North America* 21:555–560; and Colditz (1993) *Cancer Suppl.* 71:1480–1489.

5           Despite ongoing improvement in the understanding of the disease, breast cancer has remained resistant to medical intervention. Most clinical initiatives are focused on early diagnosis, followed by conventional forms of intervention, particularly surgery and chemotherapy. Such interventions are of limited success, particularly in patients where the tumor has undergone metastasis. There is a pressing need to improve the arsenal of  
10       therapies available to provide more precise and more effective treatment in a less invasive way.

          Prostate cancer is the fastest growing neoplasm in men with an estimated 244,000 new cases in the United States being diagnosed in 1995, of which approximately 44,000 deaths will result. Prostate cancer is now the most frequently  
15       diagnosed cancer in men. Prostate cancer is latent; many men carry prostate cancer cells without overt signs of disease. It is associated with a high morbidity. Cancer metastasis to bone (late stage) is common and is almost always fatal.

          Current treatments include radical prostatectomy, radiation therapy, hormonal ablation and chemotherapy. Unfortunately, in approximately 80% of cases, diagnosis  
20       of prostate cancer is established when the disease has already metastasized to the bones, thus limiting the effectiveness of surgical treatments. Hormonal therapy frequently fails with time with the development of hormone-resistant tumor cells. Although chemotherapeutic agents have been used in the treatment of prostate cancer, no single agent has demonstrated superiority over its counterparts, and no drug  
25       combination seems particularly effective. The generally drug-resistant, slow-growing nature of most prostate cancers makes them particularly unresponsive to standard chemotherapy.

A major, indeed the overwhelming, obstacle to cancer therapy is the problem of selectivity; that is, the ability to inhibit the multiplication of tumor cells, while leaving unaffected the function of normal cells. For example, in prostate cancer therapy, the therapeutic ratio, or ratio of tumor cell killing to normal cell killing of traditional tumor chemotherapy, is only 1.5:1. Thus, more effective treatment methods and pharmaceutical compositions for therapy and prophylaxis of neoplasia are needed.

Solid tumors frequently contain regions that are poorly vascularized, partly because the tumor cells grow faster than the endothelial cells that make up the blood vessels. Tumor cells can remain viable in such hypoxic conditions and are often refractory to chemotherapy and radiation therapy. In a recent study of cervical cancer, the oxygen status of a tumor was shown to be the single most important prognostic factor, ahead of age of patient, menopausal status, clinical stage, size and histology. Hoeckel et al. (1996) *Semin. Radiat. Oncol.* 6:1-8.

Of particular interest is development of more specific, targeted forms of cancer therapy, especially for cancers that are difficult to treat successfully. In contrast to conventional cancer therapies, which result in relatively non-specific and often serious toxicity, more specific treatment modalities attempt to inhibit or kill malignant cells selectively while leaving healthy cells intact. Radioresistant and chemoresistant tumors present particular challenges, and there is a need for methods of treating these types of tumors.

One possible treatment approach for many of these cancers is gene therapy, whereby a gene of interest is introduced into the malignant cell. Various viral vectors, including adenoviral vectors, have been developed as vehicles for gene therapy. The virtually exclusive focus in development of adenoviral vectors for gene therapy is use of adenovirus merely as a vehicle for introducing the gene of interest, not as an effector in itself. Replication of adenovirus has been viewed as an undesirable result, largely due to the host immune response. In the treatment of

cancer by replication-defective adenoviruses, the host immune response limits the duration of repeat doses at two levels. First, the capsid proteins of the adenovirus delivery vehicle itself are immunogenic. Second, viral late genes are frequently expressed in transduced cells, eliciting cellular immunity. Thus, the ability to repeatedly administer cytokines, tumor suppressor genes, ribozymes, suicide genes, or genes which convert prodrug to an active drug has been limited by the immunogenicity of both the gene transfer vehicle and the viral gene products of the transfer vehicle as well as the transient nature of gene expression.

Use of adenoviral vectors as therapeutic vehicles for cancer has been reported. Some of these approaches utilize tissue (i.e., cell type) specific transcriptional regulatory elements to selectively drive adenoviral replication (and thus cytotoxicity). U.S. Pat. No. 5,698,443; see also WO 95/11984; WO 96/17053; WO 96/34969; WO 98/35028. While useful and promising, there remain other treatment contexts for which tissue specific replication may be insufficient.

Besides cancerous cells, it is often desirable to selectively destroy certain unwanted cells or tissues. Besides surgery, however, which is invasive, there is a dearth of methods available, particularly non-invasive methods, which would allow such selective cytotoxicity and/or suppression.

There is a need for vector constructs that are capable of eliminating essentially all cancerous cells in a minimum number of administrations before specific immunological response against the vector prevents further treatment and which are suitable for use in specific, focused cancer ablation treatments. There is also a need for an ability to selectively destroy, or impair, unwanted cells, regardless of cell type and/or regardless of anatomical location.

## SUMMARY OF THE INVENTION

Replication-competent adenoviral vectors specific for cells in a given, or particular, physiological state that permits or induces expression of polynucleotides under transcriptional control of a cell status-specific TRE, and methods for their use are provided. In these replication-competent adenovirus vectors, one or more adenoviral genes is under transcriptional control of an cell status-specific transcriptional regulatory element (TRE). Preferably, the adenoviral gene under transcriptional control of a cell status-specific TRE is one that is essential for adenoviral propagation. A transgene under control of the cell status-specific TRE may also be present. For the adenoviral vectors of the present invention, a cell status-specific TRE is active in a cell existing in a particular, reversible, physiological state, which may be an aberrant physiological state, i.e., a physiological state that deviates from the typical, or normal, physiological state of that same cell when in a non-dividing or regulated dividing state under normal, physiological conditions.

Accordingly, in one aspect, the invention provides an adenovirus vector comprising an adenovirus gene, wherein said adenovirus gene is under transcriptional control of a cell status-specific TRE. In another embodiment, a cell status-specific TRE is human. In another embodiment, a cell status-specific TRE comprises a cell status-specific promoter and enhancer. In yet another embodiment, a cell status-specific TRE is juxtaposed with a cell type-specific TRE, and together the two elements control expression of an adenovirus gene. Thus, the invention provides adenovirus vectors comprising a TRE comprising a cell status-specific TRE and a cell type-specific TRE.

In some embodiments, the adenovirus gene under transcriptional control of a cell status-specific TRE is an adenovirus gene essential for replication. In some embodiments, the adenoviral gene essential for replication is an early gene. In another embodiment, the early gene is E1A. In another embodiment, the early gene is E1B. In yet another embodiment, both E1A and E1B are under transcriptional



control of a cell status-specific TRE. In other embodiments, the adenovirus gene essential for replication is a late gene.

In another embodiment, the cell status-specific TRE comprises a hypoxia responsive element. In another embodiment, the cell status-specific TRE comprises the nucleotide sequence of SEQ ID NO:1.

In another embodiment, the cell status-specific TRE comprises a cell cycle-specific TRE. The cell cycle-specific TRE can be derived from the E2F1 5' flanking region. In one embodiment, the cell cycle-specific TRE comprises the nucleotide sequence depicted in SEQ ID NO:2.

In other embodiments, the adenovirus vector can further comprise a transgene, wherein said transgene is under transcriptional control of an cell status-specific TRE. In some embodiments, the transgene is a cytotoxic gene.

In other embodiments, the adenoviral vector comprises an adenoviral gene essential for adenoviral replication under control of a first cell status-specific TRE, and a second adenoviral gene essential for adenoviral replication under control of a second cell status-specific TRE. The first and the second cell status-specific TREs can be identical, substantially identical, or different from, one another.

In other embodiments, the adenoviral vector comprises an adenoviral gene essential for adenoviral replication under control of a first cell status-specific TRE, and a transgene under control of a second cell status-specific TRE. The first and the second cell status-specific TREs can be substantially identical to, or different from, one another.

In other embodiments, the adenovirus vector comprises an adenovirus gene under transcriptional control of a cell status-specific TRE, and a second adenovirus gene under transcriptional control of a cell type-specific TRE. In other embodiments, the adenovirus vector comprises an adenovirus gene under transcriptional control of a cell status-specific TRE, and a transgene under transcriptional control of a cell type-specific TRE.

In another aspect, the invention provides a host cell comprising the adenovirus vector(s) described herein.

In another aspect, the invention provides pharmaceutical compositions comprising an adenovirus vector(s) described herein.

5 In another aspect, the invention provides kits which contain an adenoviral vector(s) described herein.

In another aspect, methods are provided for conferring selective cytotoxicity in target cells (i.e., cells which permit or induce a cell status-specific TRE to function), comprising contacting the cells with an adenovirus vector(s) described herein,  
10 whereby the vector enters the cell.

Another embodiment of the invention is an adenovirus which replicates preferentially in mammalian cells whose cell status permits or induces the function of a cell status-specific TRE.

In another aspect, methods are provided for propagating an adenovirus specific for mammalian cells whose cell status permits the function of a cell status-specific TRE, said method comprising combining an adenovirus vector(s) described  
15 herein with mammalian cells whose cell status permits the function of a cell status-specific TRE, whereby said adenovirus is propagated.

The invention further provides methods of suppressing tumor cell growth, more particularly a target tumor cell (i.e., a tumor cell that permits or induces a cell status-specific TRE to function), comprising contacting a tumor cell with an  
20 adenoviral vector of the invention such that the adenoviral vector enters the tumor cell and exhibits selective cytotoxicity for the tumor cell.

In another aspect, methods are provided for detecting cells whose cell status  
25 permits the function of a cell status-specific TRE in a biological sample, comprising contacting cells of a biological sample with an adenovirus vector(s) described herein, and detecting replication of the adenovirus vector, if any.

## BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a schematic representation of adenovirus vector CN796, in which the E1A gene is under transcriptional control of an HRE and a PSA-TRE, as described in Example 1.

Figure 2 shows the nucleotide sequence of an HRE from the 5' flanking region of a rat enolase-1 gene (SEQ ID NO:1).

Figure 3 shows the nucleotide sequence of the 5' flanking region of a human E2F1 gene (SEQ ID NO:2). The asterisk indicates the transcription start site.

Figure 4 depicts a nucleotide sequence of a prostate-specific antigen TRE.

Figure 5 depicts a nucleotide sequence of a carcinoembryonic antigen TRE.

Figure 6 depicts a nucleotide sequence of a human glandular kallikrein TRE.

Figure 7 depicts a nucleotide sequence of a mucin TRE.

Figure 8 depicts a nucleotide sequence of a rat probasin TRE.

Figure 9 depicts a nucleotide sequence and translated amino acid sequence of an adenovirus death protein.

## MODES FOR CARRYING OUT THE INVENTION

We have discovered and constructed replication-competent adenovirus vectors which contain an adenoviral gene under transcriptional control of a cell status-specific transcriptional response element (TRE) such that the adenovirus gene is transcribed preferentially in cells whose cell status permit the function of the cell status-specific TRE, and have developed methods using these adenovirus vectors. In some preferred embodiments, the adenovirus vectors of this invention comprise at least one adenovirus gene necessary for adenoviral replication, preferably at least one early gene, under the transcriptional control of a TRE specifically regulated by binding of transcriptional factor(s) and/or co-factor(s) necessary for transcription regulated by the cell status-specific TRE. By providing for cell status-specific transcription of at least one adenovirus gene required for replication, the invention

provides adenovirus vectors that can be used for specific cytotoxic effects due to selective replication and/or selective transcription. This is especially useful in the cancer context, in which targeted cell killing is desirable. This is also useful for targeted cytotoxic effects in other, non-tumor cells, when selective destruction and/or suppression of these cells is desirable. The vectors can also be useful for detecting the presence of cells whose cell status permits function of a cell status-specific TRE in, for example, an appropriate biological (such as clinical) sample. Further, the adenovirus vector(s) can optionally selectively produce one or more proteins of interest in a target cell by using a cell status-specific TRE.

We have found that adenovirus vectors of the invention replicate and/or express an adenoviral gene operably linked to a cell status-specific TRE preferentially in cells whose status permits the function of a cell status-specific TRE. In contrast to previously- described adenoviral vectors designed to replicate preferentially in specific, differentiated cell types, the adenovirus vectors of the present invention comprise regulatory elements that are not cell type-specific. Rather, they confer cell status-specific adenoviral replication and/or cell status-specific expression of an operably linked adenoviral gene and/or transgene.

The adenovirus vectors of the present invention comprise a cell status-specific TRE which is functional in a cell which exhibits a particular physiological (i.e., environmental or metabolic) characteristic which is reversible and/or progressive. The target cell may exhibit an aberrant physiological state, such as low oxygen tension, or may be subjected to an aberrant environmental condition, such as heat or ionizing radiation, in order for the cell-status TRE to function. The replication preference of these vectors is indicated by comparing the level of replication (i.e., titer) in cells in a requisite physiological state or condition (for example, an aberrant physiological state) to the level of replication in cells not exhibiting the requisite physiological state (for example, under normal physiological conditions). Thus, the invention also uses and takes advantage of what has been considered an undesirable

aspect of adenoviral vectors, namely, their replication and possibly concomitant immunogenicity. The probability of runaway infection is significantly reduced due to the cell status-specific requirements for viral replication. Without wishing to be bound by any particular theory, the inventors note that production of adenovirus proteins can serve to activate and/or stimulate the immune system, generally and/or specifically toward target cells producing adenoviral proteins, which can be an important consideration in the cancer context, where patients are often moderately to severely immunocompromised.

The adenovirus vectors of the present invention find particular utility in specific treatment regimens, in which the treatment is highly focused toward, for example, a particular cancer which might otherwise be inoperable or untreatable. An important feature of the invention is that the vectors are useful in these treatments regardless of the tissue or cell type of the cancer, and yet their cytotoxicity can be targeted to certain locations.

#### *General Techniques*

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of molecular biology (including recombinant techniques), microbiology, cell biology, biochemistry, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature, such as, "Molecular Cloning: A Laboratory Manual", second edition (Sanbrook et al., 1989); "Oligonucleotide Synthesis" (M.J. Gait, ed., 1984); "Animal Cell Culture" (R.I. Freshney, ed., 1987); "Methods in Enzymology" (Academic Press, Inc.); "Handbook of Experimental Immunology" (D.M. Wei & C.C. Blackwell, eds.); "Gene Transfer Vectors for Mammalian Cells" (J.M. Miller & M.P. Calos, eds., 1987); "Current Protocols in Molecular Biology" (F.M. Ausubel et al., eds., 1987); "PCR: The Polymerase Chain Reaction", (Mullis et al., eds., 1994); "Current Protocols in Immunology" (J.E. Coligan et al., eds., 1991).

For techniques related to adenovirus, see, *inter alia*, Felgner and Ringold (1989) *Nature* 337:387-388; Berkner and Sharp (1983) *Nucl. Acids Res.* 11:6003-6020; Graham (1984) *EMBO J.* 3:2917-2922; Bett et al. (1993) *J. Virology* 67:5911-5921; Bett et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:8802-8806.

## 5      *Definitions*

As used herein, a “transcription response element” or “transcriptional regulatory element”, or “TRE” is a polynucleotide sequence, preferably a DNA sequence, which increases transcription of an operably linked polynucleotide sequence in a host cell that allows that TRE to function. A TRE can comprise an enhancer and/or a promoter.

As used herein, the term “cell status-specific TRE” is one that confers transcriptional activation on an operably linked polynucleotide in a cell which allows a cell status-specific TRE to function, i.e., a cell which exhibits a particular physiological condition, including, but not limited to, an aberrant physiological state. “Cell status” thus refers to a given, or particular, physiological state (or condition) of a cell, which is reversible and/or progressive. The physiological state may be generated internally or externally; for example, it may be a metabolic state (such as low oxygen), or it may be generated due to heat or ionizing radiation. “Cell status” is distinct from a “cell type”, which relates to a differentiation state of a cell, which under normal conditions is irreversible. Generally (but not necessarily), as discussed herein, a cell status is embodied in an aberrant physiological state, examples of which are given below.

A “normal cell status” or “normal physiological state” is the status of a cell which exists in normal physiological conditions and which is non-dividing or divides in a regulated manner, i.e., a cell in a normal physiological state.

The terms “aberrant cell status” and “aberrant physiological state”, used interchangeably herein, intend a condition of a cell which is a response to, a result of, or is influenced by, an aberrant physiological condition. An aberrant cell status is

neither cell type-specific nor tissue type-specific. An aberrant cell status is defined in relation to a cell of the same type which is in a non-dividing/regulated dividing state and under normal physiological conditions.

“Normal physiological conditions” are known to those skilled in the art.

5 These conditions may vary, depending on a cell’s location in the body. For example, oxygen tension can vary from tissue to tissue. For *in vitro* analyses for the purposes of determining whether a TRE is responsive to deviations from normal physiological conditions, these conditions generally include an oxygen concentration of about 20% O<sub>2</sub>, and a temperature of about 37°C. “Regulated cell division” is a term well  
10 understood in the art and refers to the normal mitotic activity of a cell. Those skilled in the art understand that normal mitotic activity varies from cell type to cell type. For example, many terminally differentiated cells in tissues exhibit little or no mitotic activity, while hematopoietic cells are generally mitotically active.

An “aberrant physiological condition” or “aberrant physiological state”, as  
15 used herein, intends a condition which deviates from normal physiological conditions, and includes, but is not limited to, a physiological condition that is characterized by alterations in oxygen concentration, such as hypoxic conditions; temperatures which deviate from physiological temperatures; a condition that triggers apoptosis; radiation, including ionizing radiation and UV irradiation; de-regulated  
20 cell division, resulting for example, from a lack of, or insufficient amounts of, or inactivity of, a factor which controls cell division, such as, for example, retinoblastoma protein (Rb); variations in timing of cell cycle; infection with a pathogen; exposure to a chemical substance; or a combination of the above-listed conditions. Another example is a mutation that could, or does, exist in any cell type,  
25 i.e., its existence does not depend on, or is not related to, the differentiation state of the cell.

A “target cell”, as used herein, is one that permits or induces the function of a cell status-specific TRE such that it effects transcriptional activation of an operably

linked polynucleotide. A target cell is one which exhibits a requisite physiological (or environmental) state, which may be an aberrant physiological state. Preferably, a target cell is a mammalian cell, preferably a human cell. A target cell may or may not be neoplastic. By transcriptional activation, it is intended that transcription is increased in the target cell above the levels in a control cell (e.g., a that cell when not exhibiting a requisite physiological state (generally a normal physiological state) by at least about 2 fold, preferably at least about 5 fold, preferably at least about 10 fold, more preferably at least about 20 fold, more preferably at least about 50 fold, more preferably at least about 100 fold, more preferably at least about 200 fold, even more preferably at least about 400 fold to about 500 fold, even more preferably at least about 1000 fold. The normal levels are generally the level of activity (if any) in a cell as tested under conditions that activate the cell status-specific TRE, or the level of activity (if any) of a reporter construct lacking a cell status-specific TRE as measured in a cell exhibiting the requisite physiological condition.

A “functionally-preserved” variant of a cell status-specific TRE is a cell status-specific TRE which differs from another cell status-specific TRE, but still retains cell status cell-specific transcription activity. The difference in an cell status-specific TRE can be due to differences in linear sequence, arising from, for example, single base mutation(s), addition(s), deletion(s), and/or modification(s) of the bases. The difference can also arise from changes in the sugar(s), and/or linkage(s) between the bases of a cell status-specific TRE.

An “adenovirus vector” or “adenoviral vector” (used interchangeably) comprises a polynucleotide construct of the invention. A polynucleotide construct of this invention may be in any of several forms, including, but not limited to, DNA, DNA encapsulated in an adenovirus coat, DNA packaged in another viral or viral-like form (such as herpes simplex, and AAV), DNA encapsulated in liposomes, DNA complexed with polylysine, complexed with synthetic polycationic molecules, conjugated with transferrin, and complexed with compounds such as PEG to



immunologically “mask” the molecule and/or increase half-life, and conjugated to a nonviral protein. Preferably, the polynucleotide is DNA. As used herein, “DNA” includes not only bases A, T, C, and G, but also includes any of their analogs or modified forms of these bases, such as methylated nucleotides, internucleotide  
5 modifications such as uncharged linkages and thioates, use of sugar analogs, and modified and/or alternative backbone structures, such as polyamides. For purposes of this invention, adenovirus vectors are replication-competent in a target cell.

The terms “polynucleotide” and “nucleic acid”, used interchangeably herein, refer to a polymeric form of nucleotides of any length, either ribonucleotides or  
10 deoxyribonucleotides. These terms include a single-, double- or triple-stranded DNA, genomic DNA, cDNA, RNA, DNA-RNA hybrid, or a polymer comprising purine and pyrimidine bases, or other natural, chemically, biochemically modified, non-natural or derivatized nucleotide bases. The backbone of the polynucleotide can comprise sugars and phosphate groups (as may typically be found in RNA or DNA),  
15 or modified or substituted sugar or phosphate groups. Alternatively, the backbone of the polynucleotide can comprise a polymer of synthetic subunits such as phosphoramidates and thus can be a oligodeoxynucleoside phosphoramidate (P-NH<sub>2</sub>) or a mixed phosphoramidate- phosphodiester oligomer. Peyrottes et al. (1996) *Nucleic Acids Res.* 24: 1841-8; Chaturvedi et al. (1996) *Nucleic Acids Res.* 24: 2318-  
20 23; Schultz et al. (1996) *Nucleic Acids Res.* 24: 2966-73. A phosphorothiate linkage can be used in place of a phosphodiester linkage. Braun et al. (1988) *J. Immunol.* 141: 2084-9; Latimer et al. (1995) *Mol. Immunol.* 32: 1057-1064. In addition, a double-stranded polynucleotide can be obtained from the single stranded polynucleotide product of chemical synthesis either by synthesizing the  
25 complementary strand and annealing the strands under appropriate conditions, or by synthesizing the complementary strand de novo using a DNA polymerase with an appropriate primer.

The following are non-limiting examples of polynucleotides: a gene or gene fragment, exons, introns, mRNA, tRNA, rRNA, ribozymes, cDNA, recombinant polynucleotides, branched polynucleotides, plasmids, vectors, isolated DNA of any sequence, isolated RNA of any sequence, nucleic acid probes, and primers. A polynucleotide may comprise modified nucleotides, such as methylated nucleotides and nucleotide analogs, uracyl, other sugars and linking groups such as fluororibose and thioate, and nucleotide branches. The sequence of nucleotides may be interrupted by non-nucleotide components. A polynucleotide may be further modified after polymerization, such as by conjugation with a labeling component. Other types of modifications included in this definition are caps, substitution of one or more of the naturally occurring nucleotides with an analog, and introduction of means for attaching the polynucleotide to proteins, metal ions, labeling components, other polynucleotides, or a solid support. Preferably, the polynucleotide is DNA. As used herein, "DNA" includes not only bases A, T, C, and G, but also includes any of their analogs or modified forms of these bases, such as methylated nucleotides, internucleotide modifications such as uncharged linkages and thioates, use of sugar analogs, and modified and/or alternative backbone structures, such as polyamides.

A polynucleotide or polynucleotide region has a certain percentage (for example, 80%, 85%, 90%, or 95%) of "sequence identity" to another sequence means that, when aligned, that percentage of bases are the same in comparing the two sequences. This alignment and the percent homology or sequence identity can be determined using software programs known in the art, for example those described in *Current Protocols in Molecular Biology* (F.M. Ausubel et al., eds., 1987) Supplement 30, section 7.7.18, Table 7.7.1. A preferred alignment program is ALIGN Plus (Scientific and Educational Software, Pennsylvania), preferably using default parameters.

"Under transcriptional control" is a term well-understood in the art and indicates that transcription of a polynucleotide sequence, usually a DNA sequence,

depends on its being operably (operatively) linked to an element which contributes to the initiation of, or promotes, transcription. “Operably linked” refers to a juxtaposition wherein the elements are in an arrangement allowing them to function.

“Replication” and “propagation” are used interchangeably and refer to the ability of a polynucleotide construct of the invention to reproduce, or proliferate. This term is well understood in the art. For purposes of this invention, replication involves production of adenovirus proteins and is generally directed to reproduction of adenovirus. Replication can be measured using assays standard in the art and described herein, such as a burst assay, plaque assay, or a one-step growth curve assay.

As used herein, “cytotoxicity” is a term well understood in the art and refers to a state in which a cell’s usual biochemical or biological activities are compromised (i.e., inhibited). These activities include, but are not limited to, metabolism; cellular replication; DNA replication; transcription; translation; uptake of molecules.

“Cytotoxicity” includes cell death and/or cytolysis. Assays are known in the art which indicate cytotoxicity, such as dye exclusion, <sup>3</sup>H-thymidine uptake, and plaque assays.

The term “selective cytotoxicity”, as used herein, refers to the cytotoxicity conferred by an adenovirus vector of the present invention on a cell which allows or induces a cell status-specific TRE to function (a target cell) when compared to the cytotoxicity conferred by an adenoviral vector of the present invention on a cell which does not allow a cell status-specific TRE to function (a non-target cell). Such cytotoxicity may be measured, for example, by plaque assays, by reduction or stablization in size of a tumor comprising target cells, or the reduction or stabilization of serum levels of a marker characteristic of the tumor cells, or a tissue-specific marker, e.g., a cancer marker, such as prostate specific antigen.

In the context of adenovirus, a “heterologous polynucleotide” or “heterologous gene” or “transgene” is any polynucleotide or gene that is not present

in wild-type adenovirus. Preferably, the transgene will also not be expressed or present in the target cell prior to introduction by the adenovirus vector. Examples of preferred transgenes are provided below.

5 In the context of adenovirus, a “heterologous” promoter or enhancer is one which is not associated with or derived from an adenovirus gene.

In the context of adenovirus, an “endogenous” promoter, enhancer, or TRE is native to or derived from adenovirus.

10 In the context of a cell status-specific TRE, a “heterologous” promoter or enhancer is one which is not normally associated in a cell with or derived from a cell status-specific TRE. Examples of a heterologous promoter or enhancer are the albumin promoter or enhancer and other viral promoters and enhancers, such as SV40, or cell type-specific TREs such as a prostate-specific TRE.

15 A “cell type-specific TRE” is preferentially functional in a specific type of cell relative to other types of cells. In contrast to cell status, “cell type” is a reflection of a differentiation state of a cell which is irreversible. For example, a prostate-specific antigen is expressed in prostate cells, but is not substantially expressed in other cell types such as hepatocytes, astrocytes, cardiocytes, lymphocytes, etc. Generally, a cell type-specific TRE is active in only one cell type. When a cell type-specific TRE is active in more than one cell type, its activity is restricted to a limited number of cell types, i.e., it is not active in all cell types. A cell type-specific TRE may or may not be tumor cell specific.

20 “Suppressing” tumor growth indicates a growth state that is curtailed when compared to growth without contact with, i.e., transfection by, an adenoviral vector described herein. Tumor cell growth can be assessed by any means known in the art, including, but not limited to, measuring tumor size, determining whether tumor cells are proliferating using a  $^3\text{H}$ -thymidine incorporation assay, or counting tumor cells. “Suppressing” tumor cell growth means any or all of the following states: slowing, delaying, and stopping tumor growth, as well as tumor shrinkage.

As used herein, the terms “neoplastic cells”, “neoplasia”, “tumor”, “tumor cells”, “cancer” and “cancer cells”, (used interchangeably) refer to cells which exhibit relatively autonomous growth, so that they exhibit an aberrant growth phenotype characterized by a significant loss of control of cell proliferation (i.e., de-regulated cell division). Neoplastic cells can be malignant or benign.

A “host cell” includes an individual cell or cell culture which can be or has been a recipient of an adenoviral vector(s) of this invention. Host cells include progeny of a single host cell, and the progeny may not necessarily be completely identical (in morphology or in total DNA complement) to the original parent cell due to natural, accidental, or deliberate mutation and/or change. A host cell includes cells transfected or infected *in vivo* or *in vitro* with an adenoviral vector of this invention.

“Replication” and “propagation” are used interchangeably and refer to the ability of an adenovirus vector of the invention to reproduce or proliferate. These terms are well understood in the art. For purposes of this invention, replication involves production of adenovirus proteins and is generally directed to reproduction of adenovirus. Replication can be measured using assays standard in the art and described herein, such as a burst assay or plaque assay. “Replication” and “propagation” include any activity directly or indirectly involved in the process of virus manufacture, including, but not limited to, viral gene expression; production of viral proteins, nucleic acids or other components; packaging of viral components into complete viruses; and cell lysis.

An “ADP coding sequence” is a polynucleotide that encodes ADP or a functional fragment thereof. In the context of ADP, a “functional fragment” of ADP is one that exhibits cytotoxic activity, especially cell lysis, with respect to adenoviral replication. Ways to measure cytotoxic activity are known in the art and are described herein.

A polynucleotide that “encodes” an ADP polypeptide is one that can be transcribed and/or translated to produce an ADP polypeptide or a fragment thereof. The anti-sense strand of such a polynucleotide is also said to encode the sequence.

An “ADP polypeptide” is a polypeptide containing at least a portion, or region, of the amino acid sequence of an ADP (see, for example, SEQ ID NO:5), and which displays a function associated with ADP, particularly cytotoxicity, more particularly, cell lysis. As discussed herein, these functions can be measured using techniques known in the art. It is understood that certain sequence variations may be used, due to, for example, conservative amino acid substitutions, which may provide ADP polypeptides.

A polynucleotide sequence that is “depicted in” a SEQ ID NO means that the sequence is present as an identical contiguous sequence in the SEQ ID NO. The term encompasses portions, or regions of the SEQ ID NO as well as the entire sequence contained within the SEQ ID NO.

A “biological sample” encompasses a variety of sample types obtained from an individual and can be used in a diagnostic or monitoring assay. The definition encompasses blood and other liquid samples of biological origin, solid tissue samples such as a biopsy specimen or tissue cultures or cells derived therefrom, and the progeny thereof. The definition also includes samples that have been manipulated in any way after their procurement, such as by treatment with reagents, solubilization, or enrichment for certain components, such as proteins or polynucleotides. The term “biological sample” encompasses a clinical sample, and also includes cells in culture, cell supernatants, cell lysates, serum, plasma, biological fluid, and tissue samples.

An “individual” is a vertebrate, preferably a mammal, more preferably a human. Mammals include, but are not limited to, farm animals, sport animals, rodents, primates, and pets.

An “effective amount” is an amount sufficient to effect beneficial or desired results, which may include clinical results. An effective amount can be administered

in one or more administrations. For purposes of this invention, an effective amount of an adenoviral vector is an amount that is sufficient to palliate, ameliorate, stabilize, reverse, slow or delay the progression of the disease state.

***Adenoviral vectors comprising a cell status-specific TRE***

5           The present invention provides adenoviral vector constructs which comprise an adenovirus gene under transcriptional control of a cell status-specific TRE. Preferably, the adenovirus gene contributes to cytotoxicity (whether direct and/or indirect), more preferably one that contributes to or causes cell death, even more preferably is essential for adenoviral replication. Examples of a gene that  
10           contributes to cytotoxicity include, but are not limited to, adenovirus death protein (ADP; discussed below). When the adenovirus vector(s) is selectively (i.e., preferentially) replication competent for propagation in target cells, i.e., cells which permit or induce a cell-status TRE to function, these cells will be preferentially killed upon adenoviral proliferation. Once the target cells are destroyed due to selective  
15           cytotoxic and/or cytolytic replication, the adenovirus vector replication is significantly reduced, thus lessening the probability of runaway infection and undesirable bystander effects. *In vitro* cultures may be retained to monitor the mixture (such as, for example, a biopsy or other appropriate biological sample) for occurrence (i.e., presence) and/or recurrence of the target cell, e.g., a neoplastic cell  
20           or other undesired cell. To further ensure cytotoxicity, one or more transgenes having a cytotoxic effect may also be present and under selective transcriptional control. In this embodiment, one may provide higher confidence that the target cells will be destroyed. Additionally, or alternatively, an adenovirus gene that contributes to cytotoxicity and/or cell death (such as ADP) may be included in the adenoviral  
25           vector, either free of, or under, selective transcriptional control.

Cell status-specific TREs for use in the adenoviral vectors of the present invention can be derived from any species, preferably a mammal. A number of genes have been described which are expressed in response to, or in association with, a cell

status. Any of these cell status-associated genes may be used to generate a cell status-specific TRE.

An example of a cell status is cell cycle. An exemplary gene whose expression is associated with cell cycle is E2F-1, a ubiquitously expressed, growth-regulated gene, which exhibits peak transcriptional activity in S phase. Johnson et al. (1994) *Genes Dev.* 8:1514-1525. The RB protein, as well as other members of the RB family, form specific complexes with E2F-1, thereby inhibiting its ability to activate transcription. Thus, E2F-1-responsive promoters are down-regulated by RB. Many tumor cells have disrupted RB function, which can lead to de-repression of E2F-1-responsive promoters, and, in turn, de-regulated cell division.

Accordingly, in one embodiment, the invention provides an adenoviral vector in which an adenoviral gene (preferably a gene necessary for replication) is under transcriptional control of a cell status-specific TRE, wherein the cell status-specific TRE comprises a cell cycle-activated, or cell-cycle specific, TRE. In one embodiment, the cell cycle-activated TRE is an E2F1 TRE. In one embodiment, this TRE comprises the sequence depicted in Figure 3 and SEQ ID NO:2.

Another group of genes which are regulated by cell status are those whose expression is increased in response to hypoxic conditions. Bunn and Poyton (1996) *Physiol. Rev.* 76:839-885; Dachs and Stratford (1996) *Br. J. Cancer* 74:5126-5132; Guillemin and Krasnow (1997) *Cell* 89:9-12. Many tumors have insufficient blood supply, due in part to the fact that tumor cells typically grow faster than the endothelial cells that make up the blood vessels, resulting in areas of hypoxia in the tumor. Folkman (1989) *J. Natl. Cancer Inst.* 82:4-6; and Kallinowski (1996) *The Cancer J.* 9:37-40. An important mediator of hypoxic responses is the transcriptional complex HIF-1, or hypoxia inducible factor-1, which interacts with a hypoxia-responsive element (HRE) in the regulatory regions of several genes, including vascular endothelial growth factor, and several genes encoding glycolytic enzymes, including enolase-1. Murine HRE sequences have been identified and characterized.



Firth et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:6496-6500. An HRE from a rat enolase-1 promoter is described in Jiang et al. (1997) *Cancer Res.* 57:5328-5335. An HRE from a rat enolase-1 promoter is depicted in Figure 2 and given as SEQ ID NO:1.

5           Accordingly, in one embodiment, an adenovirus vector comprises an adenovirus gene, preferably an adenoviral gene essential for replication, under transcriptional control of a cell status-specific TRE comprising an HRE. In one embodiment, the cell status-specific TRE comprises the HRE depicted in Figure 2 and SEQ ID NO:1.

10           Other cell status-specific TREs include heat-inducible (i.e., heat shock) promoters, and promoters responsive to radiation exposure, including ionizing radiation and UV radiation. For example, the promoter region of the early growth response-1 (Egr-1) gene contains an element(s) inducible by ionizing radiation. Hallahan et al. (1995) *Nat. Med.* 1:786-791; and Tsai-Morris et al. (1988) *Nucl. Acids. Res.* 16:8835-8846. Heat-inducible promoters, including heat-inducible elements, have been described. See, for example Welsh (1990) in "Stress Proteins in Biology and Medicine", Morimoto, Tisseres, and Georgopoulos, eds. Cold Spring Harbor Laboratory Press; and Perisic et al. (1989) *Cell* 59:797-806. Accordingly, in some embodiments, the cell status-specific TRE comprises an element(s) responsive to ionizing radiation. In one embodiment, this TRE comprises a 5' flanking sequence of an Egr-1 gene. In other embodiments, the cell status-specific TRE comprises a heat shock responsive, or heat-inducible, element.

20           A cell status-specific TRE can also comprise multimers. For example, an HRE can comprise a tandem series of at least two, at least three, at least four, or at least five hypoxia-responsive elements. These multimers may also contain heterologous promoter and/or enhancer sequences.

25           A cell status-specific TRE may or may not lack a silencer. The presence of a silencer (i.e., a negative regulatory element) may assist in shutting off transcription

(and thus replication) in non-permissive cells (i.e., cell in a normal cell state). Thus, presence of a silencer may confer enhanced cell status-specific replication by more effectively preventing adenoviral vector replication in non-target cells. Alternatively, lack of a silencer may assist in effecting replication in target cells, thus conferring enhanced cell status-specific replication due to more effective replication in target cells.

In other embodiments, the adenoviral vector comprises an adenoviral gene essential for adenoviral replication under control of a first cell status-specific TRE, and a second adenoviral gene essential for adenoviral replication under control of a second cell status-specific TRE. The first and the second cell status-specific TREs may or may not be identical, and may or may not be substantially identical to one another. By “substantially identical” is meant a requisite degree of sequence identity between the two TREs. The degree of sequence identity between these TREs is at least about 80%, preferably at least about 85%, more preferably at least about 90%, even more preferably at least about 95%, even more preferably at least about 98%, and most preferably 100%. Sequence identity can be determined by a sequence comparison using, i.e., sequence alignment programs that are known in the art, such as those described in *Current Protocols in Molecular Biology* (F.M. Ausubel et al., eds., 1987) Supplement 30, section 7.7.18, Table 7.7.1. A preferred alignment program is ALIGN Plus (Scientific and Educational Software, Pennsylvania), preferably using default parameters. Alternatively, hybridization under stringent conditions can also indicate degree of sequence identity. Stringent conditions are known in the art; an example of a stringent condition is 80°C (or higher temperature) and 6 X SSC (or less concentrated SSC). Other hybridization conditions and parameters (in order of increasing stringency) are: incubation temperatures of 25°C, 37°C, 50°C, and 68°C; buffer concentrations of 10 X SSC, 6 X SSC, 1 X SSC, 0.1 X SSC (where 1 X SSC is 0.15 M NaCl and 15 mM citrate buffer) and their equivalents using other buffer systems; formamide concentrations of 0%, 25%, 50%, and 75%;

incubation times from about 24 hours about 5 minutes; 1, 2, or more washing steps; wash incubation times of 1, 2, or 15 minutes; and wash solutions of 6 X SSC, 1 X SSC, 0.1 X SSC, or deionized water.

Adenoviral constructs in which the first and second cell status-specific TREs are identical or substantially identical, particularly if these TREs control transcription of early genes (such as E1A and E1B), may display an instability which may be desirable in certain contexts, such as when an automatic “self-destruction” property can shut down the virus, thereby controlling the degree of propagation. Accordingly, in some embodiments, the first and second cell status-specific TRE, or the first and second TRE (one of which is a cell-status-specific TRE) are sufficiently identical to confer instability when compared to two TREs which are less identical with respect to each other (i.e., have more sequence divergence or dissimilarity). Preferred embodiments are those in which the two TREs control E1A and E1B respectively. “Instability” means that the structural integrity of the adenoviral vectors is not preserved as the virus replicates in cells, and can be measured using standard methods in the art, such as Southern analysis. In other embodiments, the first and second TREs are sufficiently divergent and/or placed in the vector such that the vector is stable (i.e., the structural integrity of the adenoviral vector is preserved).

In other embodiments, the adenoviral vector comprises an adenoviral gene essential for adenoviral replication under control of a first cell status-specific TRE, and a transgene under control of a second cell status-specific TRE. The first and the second cell status-specific TREs may or may not be substantially identical to one another.

In some embodiments, a cell status-specific TRE can be juxtaposed with another TRE, such as a different cell status-specific TRE, or, alternatively, a cell type-specific TRE. “Juxtaposed” means a cell status-specific TRE and the second TRE transcriptionally control the same gene, or at least are proximate with respect to the same gene. For these embodiments, the cell status-specific TRE and the second

TRE may be in any of a number of configurations, including, but not limited to, (a) next to each other (i.e., abutting); (b) both 5' to the gene that is transcriptionally controlled (i.e., may have intervening sequences between them); (c) one TRE 5' and the other TRE 3' to the gene. For example, as described in Example 1 and shown in

Figure 1, a cell type-specific TRE can be juxtaposed with a cell status-specific TRE to control transcription of an operably linked adenoviral gene. Such "composite" TREs can be used to confer both cell status- and cell type-specific expression of an operably linked polynucleotide, and thus may confer significantly greater specificity and/or efficacy. Examples of cell type-specific TREs are provided below.

Alternatively, "composite" TREs can be used to confer different, and possibly synergistic, cell status specificity. For example, a composite TRE could confer specificity to hypoxia and heat shock.

Example 1 provides a description of an adenovirus construct in which a composite TRE upstream of E1A consisting of an HRE and a prostate-specific TRE, PSA-TRE (which consists of enhancer sequences -5322 to -3738 fused to PSA promoter sequence -541 to +12; see U.S. Pat. Nos. 5,871,726; 5,648,478). Accordingly, in some embodiments, the invention provides an adenovirus vector comprising an adenovirus gene essential for replication, preferably an early gene, preferably E1A or E1B, under transcriptional control of a TRE comprising an HRE (preferably comprising or consisting of the 67-base fragment depicted in SEQ ID NO:1) and a prostate cell specific TRE, preferably comprising a PSA enhancer (preferably -5322 to -3738; or about 503 to about 2086 of SEQ ID NO:3 (bases about 503 to about 2086 of Figure 4), and a promoter, preferably comprising a PSA enhancer and a PSA promoter (about 5285 to about 5836 of SEQ ID NO:3).

As is readily appreciated by one skilled in the art, a cell status-specific TRE is a polynucleotide sequence, and, as such, can exhibit function over a variety of sequence permutations. Methods of nucleotide substitution, addition, and deletion are known in the art, and readily available functional assays (such as the CAT or

luciferase reporter gene assay) allow one of ordinary skill to determine whether a sequence variant exhibits requisite cell status-specific transcription function. Hence, the invention also includes functionally-preserved variants of the nucleic acid sequences disclosed herein, which include nucleic acid substitutions, additions, and/or deletions. While not wishing to be bound by a single theory, the inventors note that it is possible that certain modifications will result in modulated resultant expression levels, including enhanced expression levels. Achievement of modulated resultant expression levels, preferably enhanced expression levels, may be especially desirable in the case of certain, more aggressive forms of cancer, or when a more rapid and/or aggressive pattern of cell killing is warranted (due to an immunocompromised condition of the individual, for example).

As an example of how cell status-specific TRE activity can be determined, a polynucleotide sequence or set of such sequences can be generated using methods known in the art, such as chemical synthesis, site-directed mutagenesis, PCR, and/or recombinant methods. The sequence(s) to be tested is inserted into a vector containing an appropriate reporter gene, including, but not limited to, chloramphenicol acetyl transferase (CAT),  $\beta$ -galactosidase (encoded by the lacZ gene), luciferase (encoded by the luc gene), green fluorescent protein, alkaline phosphatase, and horse radish peroxidase. Such vectors and assays are readily available, from, inter alia, commercial sources. Plasmids thus constructed are transfected into a suitable host cell to test for expression of the reporter gene as controlled by the putative cell status-specific TRE using transfection methods known in the art, such as calcium phosphate precipitation, electroporation, liposomes (lipofection), and DEAE-dextran. Suitable host cells include any cell type, including but not limited to, Hep3B, Hep G2, HuH7, HuH1/C12, LNCaP, HBL-100, Chang liver cells, MCF-7, HLF, HLE, 3T3, HUVEC, and HeLa. Host cells transfected with the TRE-reporter gene construct to be tested are subjected to conditions which result in a change in cell status (for example, one which result in an aberrant physiological

state). The same cells not subjected to these conditions, i.e., which are under normal physiological conditions and therefore in a normal physiological state, serve as controls. Results are obtained by measuring the level of expression of the reporter gene using standard assays. Comparison of expression between cells in a particular state and control indicates presence or absence of transcriptional activation.

“Transcriptional activation” has been defined above.

Comparisons between or among various cell status-specific TREs can be assessed, for example, by measuring and comparing levels of expression within a single cell line under normal and aberrant physiological conditions. These comparisons may also be made by measuring and comparing levels of expression within a single cell line subjected to reversible environmental conditions (such as heat) and cells not subjected to such conditions. It is understood that absolute transcriptional activity of an cell status-specific TRE will depend on several factors, such as the nature of the target cell, delivery mode and form of the cell status-specific TRE, and the coding sequence that is to be selectively transcriptionally activated. To compensate for various plasmid sizes used, activities can be expressed as relative activity per mole of transfected plasmid. Alternatively, the level of transcription (i.e., mRNA) can be measured using standard Northern analysis and hybridization techniques. Levels of transfection (i.e., transfection efficiencies) are measured by co-transfecting a plasmid encoding a different reporter gene under control of a different TRE, such as the cytomegalovirus (CMV) immediate early promoter. This analysis can also indicate negative regulatory regions, i.e., silencers.

As an example of how hypoxia induction can be measured, one can use an assay such as that described in Jiang et al. (1997) *Cancer Research* 57:5328-5335 or Dachs et al. (1997) *Nature Med.* 3:515-520. For example, a construct comprising a putative HRE, or multiple tandem copies thereof, together with a minimal promoter element, operably linked and controlling transcription of a polynucleotide which encodes a protein which is detectable or can be used to give a detectable signal, is

introduced into host cells. The host cells are then subjected to conditions of normoxia (e.g., 20% O<sub>2</sub>), and varying degrees of hypoxia, such as 5%, 2%, 1%, 0.1%, or less, O<sub>2</sub>. The expression product of the operably linked polynucleotide (reporter gene) is then measured.

5           Alternatively a putative cell status-specific TRE can be assessed for its ability to confer adenoviral replication preference for cells exhibiting the requisite physiological state, such as heat or ionizing radiation. For this assay, constructs containing an adenovirus gene essential to replication operably linked to a putative cell status-specific TRE are transfected into cells which exhibit the requisite  
10           physiological state. Viral replication in those cells is compared, for example, to viral replication by the construct in cells under normal physiological conditions (i.e., not exhibiting the requisite physiological state).

          Any of the various serotypes of adenovirus can be used, such as Ad2, Ad5, Ad12 and Ad40. For purposes of illustration, serotype Ad5 will be exemplified  
15           herein.

          When a cell status-specific TRE is used with an adenovirus gene that is essential for propagation replication competence is preferentially achievable in the target cell expressing cell status. Preferably, the gene is an early gene, such as E1A, E1B, E2, or E4. (E3 is not essential for viral replication.) More preferably, the early  
20           gene under cell status-TRE control is E1A and/or E1B. More than one early gene can be placed under control of an cell status-specific TRE. Example 1 provides a more detailed description of such constructs.

          The E1A gene is expressed immediately after viral infection (0-2 hours) and before any other viral genes. E1A protein acts as a *trans*-acting positive-acting  
25           transcriptional regulatory factor, and is required for the expression of the other early viral genes E1B, E2, E3, E4, and the promoter-proximal major late genes. Despite the nomenclature, the promoter proximal genes driven by the major late promoter are expressed during early times after Ad5 infection. Flint (1982) *Biochem. Biophys.*

*Acta* 651:175–208; Flint (1986) *Advances Virus Research* 31:169–228; Grand (1987) *Biochem. J.* 241:25–38. In the absence of a functional E1A gene, viral infection does not proceed, because the gene products necessary for viral DNA replication are not produced. Nevins (1989) *Adv. Virus Res.* 31:35–81. The transcription start site of Ad5 E1A is at 498 and the ATG start site of the E1A protein is at 560 in the virus genome.

The E1B protein functions *in trans* and is necessary for transport of late mRNA from the nucleus to the cytoplasm. Defects in E1B expression result in poor expression of late viral proteins and an inability to shut off host cell protein synthesis. The promoter of E1B has been implicated as the defining element of difference in the host range of Ad40 and Ad5: clinically Ad40 is an enterovirus, whereas Ad5 causes acute conjunctivitis. Bailey, Mackay et al. (1993) *Virology* 193:631; Bailey et al. (1994) *Virology* 202:695-706). The E1B promoter of Ad5 consists of a single high-affinity recognition site for Spl and a TATA box.

The E2 region of adenovirus codes for proteins related to replication of the adenoviral genome, including the 72 kDa DNA-binding protein, the 80 kD precursor terminal protein and the viral DNA polymerase. The E2 region of Ad5 is transcribed in a rightward orientation from two promoters, termed E2 early and E2 late, mapping at 76.0 and 72.0 map units, respectively. While the E2 late promoter is transiently active during late stages of infection and is independent of the E1A transactivator protein, the E2 early promoter is crucial during the early phases of viral replication.

The E2 late promoter overlaps with the coding sequences of a gene encoded by the counterstrand and is therefore not amenable to genetic manipulation. However, the E2 early promoter overlaps only for a few base pairs with sequences coding for a 33 kD protein on the counterstrand. Notably, the SpeI restriction site (Ad5 position 27082) is part of the stop codon for the above mentioned 33 kD protein and conveniently separates the major E2 early transcription initiation site and TATA-binding protein site from the upstream transcription factor binding sites E2F and ATF.



Therefore, insertion of a cell status-TRE having SpeI ends into the SpeI site in the +-strand would disrupt the endogenous E2 early promoter of Ad5 and should allow cell status-restricted expression of E2 transcripts.

5 The E4 gene has a number of transcription products. The E4 region codes for two polypeptides which are responsible for stimulating the replication of viral genomic DNA and for stimulating late gene expression. The protein products of open reading frames (ORFs) 3 and 6 can both perform these functions by binding the 55kD protein from E1B and heterodimers of E2F-1 and DP-1. The ORF 6 protein requires interaction with the E1B 55kD protein for activity while the ORF 3 protein  
10 does not. In the absence of functional protein from ORF 3 and ORF 6, plaques are produced with an efficiency less than  $10^{-6}$  that of wild type virus. To further restrict viral replication to cells exhibiting a requisite physiological condition or state, E4 ORFs 1-3 can be deleted, making viral DNA replication and late gene synthesis dependent on E4 ORF 6 protein. By combining such a mutant with sequences in  
15 which the E1B region is regulated by a cell status-specific TRE, a virus can be obtained in which both the E1B function and E4 function are dependent on a cell status-specific TRE driving E1B.

The major late genes relevant to the subject invention are genes L1, L2, L3, L4, and L5 which encode proteins of the adenovirus virion. All of these genes  
20 (typically coding for structural proteins) are probably required for adenoviral replication. The late genes are all under the control of the major late promoter (MLP), which is located in Ad5 at +5986 to +6048.

In addition to conferring selective cytotoxic and/or cytolytic activity by virtue of preferential replication competence in cells exhibiting a requisite physiological  
25 state (for example, an aberrant physiological state such as low oxygen conditions), the adenovirus vectors of this invention can further include a heterologous gene (transgene) under the control of a cell status-specific TRE. In this way, various genetic capabilities may be introduced into target cells, particularly cancer cells. For

example, in certain instances, it may be desirable to enhance the degree and/or rate of cytotoxic activity, due to, for example, the relatively refractory nature or particular aggressiveness of the cancerous target cell. This could be accomplished by coupling the cell status-specific replicative cytotoxic activity with cell-specific expression of, for example, HSV-tk and/or cytosine deaminase (cd), which renders cells capable of metabolizing 5-fluorocytosine (5-FC) to the chemotherapeutic agent 5-fluorouracil (5-FU). Using these types of transgenes may also confer a bystander effect.

Other desirable transgenes that may be introduced via an adenovirus vector(s) include genes encoding cytotoxic proteins, such as the A chains of diphtheria toxin, ricin or abrin (Palmiter et al. (1987) *Cell* 50: 435; Maxwell et al. (1987) *Mol. Cell. Biol.* 7: 1576; Behringer et al. (1988) *Genes Dev.* 2: 453; Messing et al. (1992) *Neuron* 8: 507; Piatak et al. (1988) *J. Biol. Chem.* 263: 4937; Lamb et al. (1985) *Eur. J. Biochem.* 148: 265; Frankel et al. (1989) *Mol. Cell. Biol.* 9: 415), genes encoding a factor capable of initiating apoptosis, sequences encoding antisense transcripts or ribozymes, which among other capabilities may be directed to mRNAs encoding proteins essential for proliferation, such as structural proteins, or transcription factors; viral or other pathogenic proteins, where the pathogen proliferates intracellularly; genes that encode an engineered cytoplasmic variant of a nuclease (e.g. RNase A) or protease (e.g. austin, papain, proteinase K, carboxypeptidase, etc.), or encode the Fas gene, and the like. Other genes of interest include cytokines, antigens, transmembrane proteins, and the like, such as IL-1, -2, -6, -12, GM-CSF, G-CSF, M-CSF, IFN- $\alpha$ , - $\beta$ , - $\gamma$ , TNF- $\alpha$ , - $\beta$ , TGF- $\alpha$ , - $\beta$ , NGF, and the like. The positive effector genes could be used in an earlier phase, followed by cytotoxic activity due to replication.

In one embodiment, the adenovirus death protein (ADP), encoded within the E3 region, is maintained in the adenovirus vector. The ADP gene, under control of the major late promoter (MLP), appears to code for a protein (ADP) that is important in expediting host cell lysis. Tollefson et al. (1996) *J. Virol.* 70(4):2296; Tollefson et

al. (1992) *J. Virol.* 66(6):3633. Thus, adenoviral vectors containing the ADP gene may render the adenoviral vector more potent, making possible more effective treatment and/or a lower dosage requirement.

Accordingly, the invention provides an adenoviral vector as described herein that further includes a polynucleotide sequence encoding an ADP. A DNA sequence encoding an ADP and the amino acid sequence of an ADP are depicted Figure 9.

Briefly, an ADP coding sequence is obtained preferably from Ad2 (since this is the strain in which ADP has been more fully characterized) using techniques known in the art, such as PCR. Preferably, the Y leader (which is an important sequence for correct expression of late genes) is also obtained and ligated to the ADP coding sequence. The ADP coding sequence (with or without the Y leader) can then be introduced into the adenoviral genome, for example, in the E3 region (where the ADP coding sequence will be driven by the MLP). The ADP coding sequence could also be inserted in other locations of the adenovirus genome, such as the E4 region.

Alternatively, the ADP coding sequence could be operably linked to a heterologous promoter (with or without enhancer(s)), including, but not limited to, another viral promoter, a cell status-specific TRE such as a hypoxia responsive element, or a cell type-specific TRE such as those derived from carcinoembryonic antigen (CEA), mucin, and rat probasin genes.

*Adenoviral vectors of the invention further comprising a cell type specific element*

In addition to conferring selective cytotoxic and/or cytolytic activity by virtue of preferential replication competence and/or by preferential transcription of a gene encoding a cytotoxic factor in cells exhibiting a requisite physiological state, the adenovirus vectors of this invention can further include an adenovirus gene and/or a heterologous gene (transgene) under the control of a cell type-specific TRE. In this way, cytotoxicity is further limited to a particular cell type.

For example, TREs that function preferentially in prostate cells include, but are not limited to, TREs derived from the prostate-specific antigen gene (*PSA*-TRE) (U.S. Patent No. 5,648,478), the glandular kallikrein-1 gene (from the human gene, *hKLK2*-TRE), and the probasin gene (*PB*-TRE) (International Patent Application No. PCT/US98/04132). All three of these genes are preferentially expressed in prostate cells and the expression is androgen-inducible. Generally, expression of genes responsive to androgen induction requires the presence of an androgen receptor (AR).

PSA is synthesized exclusively by normal, hyperplastic, and malignant prostatic epithelia; hence, its tissue-specific expression has made it an excellent biomarker for benign prostatic hyperplasia (BPH) and prostatic carcinoma (CaP). Normal serum levels of PSA are typically below 5 ng/ml, with elevated levels indicative of BPH or CaP. Lundwall et al. (1987) *FEBS Lett.* 214: 317; Lundwall (1989) *Biochem. Biophys. Res. Comm.* 161: 1151; and Riegmann et al. (1991) *Molec. Endocrin.* 5: 1921.

The region of the *PSA* gene that is used to provide cell specificity dependent upon androgens, particular in prostate cells, involves approximately 6.0 kilobases. Schuur et al. (1996) *J. Biol. Chem.* 271:7043-7051. An enhancer region of approximately 1.5 kb in humans is located between nt -5322 and nt -3738, relative to the transcription start site of the *PSA* gene. The *PSA* promoter consists of the sequence from about nt -540 to nt +12 relative to the transcription start site. Juxtapositioning of these two genetic elements yield a fully functional, minimal prostate-specific enhancer/promoter (*PSE*) TRE. Other portions of the approximately 6.0 kb region of the *PSA* gene can be used in the present invention, as long as requisite functionality is maintained. In Example 1, adenoviral vector CN796 is described which comprises a composite TRE comprising an HRE and a *PSA*-TRE, the *PSA*-TRE comprising a *PSA* enhancer from -5322 to -3738 fused to a *PSA* promoter from -541 to +12. This *PSA*-TRE is derived from adenoviral vector CN706. Rodriguez et al. (1997) *Cancer Research* 57:2559-2563. Accordingly, in

one embodiment an adenoviral vector comprises and adenovirus E1A gene under transcriptional control of a composite TRE comprising the cell status-specific TRE, HRE, and a cell type-specific TRE, a PSA-TRE.

5 The *PSE* and *PSA* TRE used in the present invention are derived from sequences depicted in Figure 4 (SEQ ID NO:3). The enhancer element is nucleotides about 503 to about 2086 of Figure 4 (SEQ ID NO:3). The promoter is nucleotides about 5285 to about 5836 of Figure 4 (SEQ ID NO:3). Accordingly, in some embodiments, the composite TRE comprises an HRE comprising SEQ ID NO:1 and a PSA-TRE comprises nucleotides about 503 to about 2086 of SEQ ID NO:3. In other  
10 embodiments, the composite TRE comprises an HRE comprising SEQ ID NO:1 and a PSA-TRE comprises nucleotides about 503 to about 2086 of SEQ ID NO:3 and nucleotides about 5285 to about 5836 of SEQ ID NO:3. As described above, these composite (HRE/PSA) TREs may be operably linked to an adenovirus gene essential for replication, especially an early gene such as E1A or E1B. Example 1 describes  
15 such a construct.

In the present invention, replication-competent adenovirus vectors comprising a cell status-specific TRE and a cell type-specific TRE may employ cell type-specific TREs that are preferentially functional in particular tumor cells. Non-limiting examples of tumor cell-specific TREs, and non-limiting examples of respective  
20 potential target cells, include TREs from the following genes:  $\alpha$ -fetoprotein (*AFP*) (liver cancer), mucin-like glycoprotein DF3 (*MUC1*) (breast carcinoma), carcinoembryonic antigen (*CEA*) (colorectal, gastric, pancreatic, breast, and lung cancers), plasminogen activator urokinase (*uPA*) and its receptor gene (breast, colon, and liver cancers), *HER-2/neu* (*c-erbB2/neu*) (breast, ovarian, stomach, and lung  
25 cancers).

Other cell type-specific TREs may be derived from the following exemplary genes (cell type in which the TREs are specifically functional are in parentheses): vascular endothelial growth factor receptor (endothelium), albumin (liver), factor VII

(liver), fatty acid synthase (liver), von Willebrand factor (brain endothelium), alpha-actin and myosin heavy chain (both in smooth muscle), synthetase I (small intestine), Na-K-Cl transporter (kidney). Additional cell type-specific TREs are known in the art.

5 AFP is an oncofetal protein, the expression of which is primarily restricted to developing tissues of endodermal origin (yolk sac, fetal liver, and gut), although the level of its expression varies greatly depending on the tissue and the developmental stage. AFP is of clinical interest because the serum concentration of AFP is elevated in a majority of hepatoma patients, with high levels of AFP found in patients with  
10 advanced disease. The serum AFP levels in patients appear to be regulated by AFP expression in hepatocellular carcinoma but not in surrounding normal liver. Thus, the AFP gene appears to be regulated to hepatoma cell-specific expression.

Cell type-specific TREs from the *AFP* gene have been identified. For example, the cloning and characterization of human AFP-specific enhancer activity is described in Watanabe et al. (1987) *J. Biol. Chem.* 262:4812-4818. The entire 5' *AFP* flanking region (containing the promoter, putative silencer, and enhancer  
15 elements) is contained within approximately 5 kb upstream from the transcription start site.

The *AFP* enhancer region in human is located between about nt -3954 and about nt -3335, relative to the transcription start site of the *AFP* gene. The human *AFP* promoter encompasses a region from about nt -174 to about nt +29.  
20

Juxtapositioning of these two genetic elements yields a fully functional *AFP*-TRE. Ido et al. (1995) describe a 259 bp promoter fragment (nt -230 to nt +29) that is specific for HCC. *Cancer Res.* 55:3105-3109. The *AFP* enhancer contains two  
25 regions, denoted A and B, located between nt -3954 and nt -3335 relative to the transcription start site. The promoter region contains typical TATA and CAAT boxes. Preferably, the *AFP*-TRE contains at least one enhancer region. More preferably, the *AFP*-TRE contains both enhancer regions.

Suitable target cells for adenoviral vectors containing *AFP*-TREs are any cell type that allow an *AFP*-TRE to function. Preferred are cells that express, or produce, AFP, including, but not limited to, tumor cells expressing AFP. Examples of such cells are hepatocellular carcinoma cells, gonadal and other germ cell tumors (especially endodermal sinus tumors), brain tumor cells, ovarian tumor cells, acinar cell carcinoma of the pancreas (Kawamoto et al. (1992) *Hepatogastroenterology* 39:282–286), primary gall bladder tumor (Katsuragi et al. (1989) *Rinsko Hoshasen* 34:371–374), uterine endometrial adenocarcinoma cells (Koyama et al. (1996) *Jpn. J. Cancer Res.* 87:612-617), and any metastases of the foregoing (which can occur in lung, adrenal gland, bone marrow, and/or spleen). In some cases, metastatic disease to the liver from certain pancreatic and stomach cancers produce AFP. Especially preferred are hepatocellular carcinoma cells and any of their metastases. AFP production can be measured using assays standard in the art, such as RIA, ELISA or Western blots (immunoassays) to determine levels of AFP protein production or Northern blots to determine levels of AFP mRNA production. Alternatively, such cells can be identified and/or characterized by their ability to activate transcriptionally an *AFP*-TRE (i.e., allow an *AFP*-TRE to function).

The protein urokinase plasminogen activator (uPA) and its cell surface receptor, urokinase plasminogen activator receptor (uPAR), are expressed in many of the most frequently occurring neoplasia and appear to represent important proteins in cancer metastasis. Both proteins are implicated in breast, colon, prostate, liver, renal, lung and ovarian cancer. Transcriptional regulatory elements that regulate uPA and uPAR transcription have been extensively studied. Riccio et al. (1985) *Nucleic Acids Res.* 13:2759-2771; Cannio et al. (1991) *Nucleic Acids Res.* 19:2303–2308.

CEA is a 180,000-Dalton glycoprotein tumor-associated antigen present on endodermally-derived neoplasia of the gastrointestinal tract, such as colorectal, gastric (stomach) and pancreatic cancer, as well as other adenocarcinomas such as breast and lung cancers. CEA is of clinical interest because circulating CEA can be

detected in the great majority of patients with CEA-positive tumors. In lung cancer, about 50% of total cases have circulating CEA, with high concentrations of CEA (greater than 20 ng/ml) often detected in adenocarcinomas. Approximately 50% of patients with gastric carcinoma are serologically positive for CEA.

5           The 5' upstream flanking sequence of the *CEA* gene has been shown to confer cell-specific activity. The *CEA* promoter region, approximately the first 424 nucleotides upstream of the translational start site in the 5' flanking region of the gene, was shown to confer cell-specific activity when the region provided higher promoter activity in CEA-producing cells than in non-producing HeLa cells..

10       Schrewe et al. (1990) *Mol. Cell. Biol.* 10:2738-2748. In addition, cell-specific enhancer regions have been found. WO/95/14100. The entire 5' *CEA* flanking region (containing the promoter, putative silencer, and enhancer elements) appears to be contained within approximately 14.5 kb upstream from the transcription start site. Richards et al. (1995); WO 95/14100. Further characterization of the 5' flanking

15       region of the *CEA* gene by Richards et al. (1995) indicated two upstream regions, -13.6 to -10.7 kb or -6.1 to -4.0 kb, when linked to the multimerized promoter resulted in high-level and selective expression of a reporter construct in CEA-producing LoVo and SW1463 cells. Richards et al. (1995) also localized the promoter region to nt -90 and nt +69 relative to the transcriptional start site, with region nt -41 to nt -18 as

20       essential for expression. WO95/14100 describes a series of 5' flanking *CEA* fragments which confer cell-specific activity, such as about nt -299 to about nt +69; about nt -90 to about nt +69; nt -14,500 to nt -10,600; nt -13,600 to nt -10,600, nt -6100 to nt -3800. In addition, cell specific transcription activity is conferred on an operably linked gene by the *CEA* fragment from nt -402 to nt +69, depicted in (SEQ

25       ID NO:6). Any *CEA*-TREs used in the present invention are derived from mammalian cells, including but not limited to, human cells. Thus, any of the *CEA*-TREs may be used in the invention as long as requisite desired functionality is displayed in the adenovirus vector. The cloning and characterization of *CEA*



sequences have been described in the literature and are thus made available for practice of this invention and need not be described in detail herein.

The protein product of the *MUC1* gene (known as mucin or MUC1 protein; episialin; polymorphic epithelial mucin or PEM; EMA; DF3 antigen; NPGP; PAS-O; or CA15.3 antigen) is normally expressed mainly at the apical surface of epithelial cells lining the glands or ducts of the stomach, pancreas, lungs, trachea, kidney, uterus, salivary glands, and mammary glands. Zotter et al. (1988) *Cancer Rev.* 11-12: 55-101; and Girling et al. (1989) *Int. J. Cancer* 43: 1072-1076. However, mucin is overexpressed in 75-90% of human breast carcinomas. Kufe et al. (1984) *Hybridoma* 3: 223-232. For reviews, see Hilkens (1988) *Cancer Rev.* 11-12: 25-54; and Taylor-Papadimitriou, et al. (1990) *J. Nucl. Med. Allied Sci.* 34: 144-150. Mucin protein expression correlates with the degree of breast tumor differentiation. Lundy et al. (1985) *Breast Cancer Res. Treat.* 5: 269-276. This overexpression appears to be controlled at the transcriptional level.

Overexpression of the *MUC1* gene in human breast carcinoma cells MCF-7 and ZR-75-1 appears to be regulated at the transcriptional level. Kufe et al. (1984); Kovarik (1993) *J. Biol. Chem.* 268:9917-9926; and Abe et al. (1990) *J. Cell. Physiol.* 143: 226-231. The regulatory sequences of the *MUC1* gene have been cloned, including the approximately 0.9 kb upstream of the transcription start site which contains a TRE that appears to be involved in cell-specific transcription. Abe et al. (1993) *Proc. Natl. Acad. Sci. USA* 90: 282-286; Kovarik et al. (1993); and Kovarik et al. (1996) *J. Biol. Chem.* 271:18140-18147.

Any *MUC1*-TREs used in the present invention are derived from mammalian cells, including but not limited to, human cells. Preferably, the *MUC1*-TRE is human. In one embodiment, the *MUC1*-TRE may contain the entire 0.9 kb 5' flanking sequence of the *MUC1* gene. In other embodiments, the *MUC1*-TREs comprise the following sequences (relative to the transcription start site of the *MUC1*

gene): about nt -725 to about nt +31, nt -743 to about nt +33, nt -750 to about nt +33, and nt -598 to about nt +485 (operably-linked to a promoter).

The *c-erbB2/neu* gene (*HER-2/neu* or *HER*) is a transforming gene that encodes a 185 kD epidermal growth factor receptor-related transmembrane glycoprotein. In humans, the *c-erbB2/neu* protein is expressed during fetal development, however, in adults, the protein is weakly detectable (by immunohistochemistry) in the epithelium of many normal tissues. Amplification and/or over-expression of the *c-erbB2/neu* gene has been associated with many human cancers, including breast, ovarian, uterine, prostate, stomach and lung cancers. The clinical consequences of the *c-erbB2/neu* protein over-expression have been best studied in breast and ovarian cancer. *c-erbB2/neu* protein over-expression occurs in 20 to 40% of intraductal carcinomas of the breast and 30% of ovarian cancers, and is associated with a poor prognosis in subcategories of both diseases. Human, rat and mouse *c-erbB2/neu* TREs have been identified and shown to confer *c-erbB2/neu* expressing cell specific activity. Tal et al. (1987) *Mol. Cell. Biol.* 7:2597–2601; Hudson et al. (1990) *J. Biol. Chem.* 265:4389–4393; Grooteclaes et al. (1994) *Cancer Res.* 54:4193–4199; Ishii et al. (1987) *Proc. Natl. Acad. Sci. USA* 84:4374–4378; Scott et al. (1994) *J. Biol. Chem.* 269:19848–19858.

The cell type-specific TREs listed above are provided as non-limiting examples of TREs that would function in the instant invention. Additional cell-specific TREs are known in the art, as are methods to identify and test cell specificity of suspected TREs.

#### *Using the adenoviral vectors of the invention*

The adenoviral vectors can be used in a variety of forms, including, but not limited to, naked polynucleotide (usually DNA) constructs; polynucleotide constructs complexed with agents to facilitate entry into cells, such as cationic liposomes or other cationic compounds such as polylysine; packaged into infectious adenovirus particles (which may render the adenoviral vector(s) more immunogenic); packaged

into other particulate viral forms such as HSV or AAV; complexed with agents (such as PEG) to enhance or dampen an immune response; complexed with agents that facilitate *in vivo* transfection, such as DOTMA™, DOTAP™, and polyamines. Thus, the invention also provides an adenovirus capable of replicating preferentially in cell status-producing cells. “Replicating preferentially” means that the adenovirus replicates more in cell exhibiting a requisite physiological state than a cell not exhibiting that state. Preferably, the adenovirus replicates at least about 2-fold higher, preferably at least about 5-fold higher, more preferably at least about 10-fold higher, still more preferably at least about 50-fold higher, even more preferably at least about 100-fold higher, still more preferably at least about 400-fold to about 500-fold higher, still more preferably at least about 1000-fold higher, most preferably at least about  $1 \times 10^6$  higher. Most preferably, the adenovirus replicates solely in cells exhibiting a requisite physiological state (that is, does not replicate or replicates at very low levels in cells not exhibiting the requisite physiological state).

If an adenoviral vector is packaged into an adenovirus, the adenovirus itself may also be selected to further enhance targeting. For example, adenovirus fibers mediate primary contact with cellular receptor(s) aiding in tropism. See, e.g., Amberg et al. (1997) *Virol.* 227:239-244. If a particular subgenus of an adenovirus serotype displayed tropism for a target cell type and/or reduced affinity for non-target cell types, such subgenus (or subgenera) could be used to further increase cell-specificity of cytotoxicity and/or cytolysis.

The adenoviral vectors may be delivered to the target cell in a variety of ways, including, but not limited to, liposomes, general transfection methods that are well known in the art (such as calcium phosphate precipitation or electroporation), direct injection, and intravenous infusion. The means of delivery will depend in large part on the particular adenoviral vector (including its form) as well as the type and location of the target cells (i.e., whether the cells are *in vitro* or *in vivo*).

If used as a packaged adenovirus, adenovirus vectors may be administered in an appropriate physiologically acceptable carrier at a dose of about  $10^4$  to about  $10^{14}$ . The multiplicity of infection will generally be in the range of about 0.001 to 100. If administered as a polynucleotide construct (i.e., not packaged as a virus) about 0.01  $\mu\text{g}$  to about 1000  $\mu\text{g}$  of an adenoviral vector can be administered. The adenoviral vector(s) may be administered one or more times, depending upon the intended use and the immune response potential of the host, and may also be administered as multiple, simultaneous injections. If an immune response is undesirable, the immune response may be diminished by employing a variety of immunosuppressants, so as to permit repetitive administration, without a strong immune response. If packaged as another viral form, such as HSV, an amount to be administered is based on standard knowledge about that particular virus (which is readily obtainable from, for example, published literature) and can be determined empirically.

#### *Host cells comprising the adenoviral vectors of the invention*

The present invention also provides host cells comprising (i.e., transformed with) the adenoviral vectors described herein. Both prokaryotic and eukaryotic host cells can be used as long as sequences requisite for maintenance in that host, such as appropriate replication origin(s), are present. For convenience, selectable markers are also provided. Prokaryotic host cells include bacterial cells, for example, *E. coli* and mycobacteria. Among eukaryotic host cells are yeast, insect, avian, plant and mammalian. Host systems are known in the art and need not be described in detail herein.

#### *Compositions of the invention*

The present invention also provides compositions, including pharmaceutical compositions, containing the adenoviral vectors described herein. Such compositions (especially pharmaceutical compositions) are useful for administration *in vivo*, for example, when measuring the degree of transduction and/or effectiveness of cell

5 killing in an individual. Pharmaceutical compositions, comprised an adenoviral vector of this invention in a pharmaceutically acceptable excipient (generally an effective amount of the adenoviral vector), are suitable for systemic administration to individuals in unit dosage forms, sterile parenteral solutions or suspensions, sterile non-parenteral solutions or oral solutions or suspensions, oil in water or water in oil emulsions and the like. Formulations for parenteral and nonparenteral drug delivery are known in the art and are set forth in *Remington's Pharmaceutical Sciences*, 19th Edition, Mack Publishing (1995). Pharmaceutical compositions also include lyophilized and/or reconstituted forms of the adenoviral vectors (including those packaged as a virus, such as adenovirus) of the invention.

10 Other compositions are used, and are useful for, detection methods described herein. For these compositions, the adenoviral vector usually is suspended in an appropriate solvent or solution, such as a buffer system. Such solvent systems are well known in the art.

15 *Kits of the invention*

The present invention also encompasses kits containing an adenoviral vector(s) of this invention. These kits can be used for diagnostic and/or monitoring purposes, preferably monitoring. Procedures using these kits can be performed by clinical laboratories, experimental laboratories, medical practitioners, or private individuals. Kits embodied by this invention allow someone to detect the presence of cell status-producing cells in a suitable biological sample, such as biopsy specimens.

20 The kits of the invention comprise an adenoviral vector described herein in suitable packaging. The kit may optionally provide additional components that are useful in the procedure, including, but not limited to, buffers, developing reagents, labels, reacting surfaces, means for detection, control samples, instructions, and interpretive information.

*Preparation of the adenovirus vectors of the invention*

The adenovirus vectors of this invention can be prepared using recombinant techniques that are standard in the art. Generally, a cell status-specific TRE is inserted 5' to the adenoviral gene of interest, preferably one or more early genes (although late gene(s) may be used). A cell status-specific TRE can be prepared using oligonucleotide synthesis (if the sequence is known) or recombinant methods (such as PCR and/or restriction enzymes). Convenient restriction sites, either in the natural adeno-DNA sequence or introduced by methods such as oligonucleotide directed mutagenesis and PCR, provide an insertion site for a cell status-specific TRE. Accordingly, convenient restriction sites for annealing (i.e., inserting) a cell status-specific TRE can be engineered onto the 5' and 3' ends of a cell status-specific TRE using standard recombinant methods, such as PCR.

Polynucleotides used for making adenoviral vectors of this invention may be obtained using standard methods in the art, such as chemical synthesis, by recombinant methods, and/or by obtaining the desired sequence(s) from biological sources.

Adenoviral vectors are conveniently prepared by employing two plasmids, one plasmid providing for the left hand region of adenovirus and the other plasmid providing for the right hand region, where the two plasmids share at least about 500 nt of middle region for homologous recombination. In this way, each plasmid, as desired, may be independently manipulated, followed by cotransfection in a competent host, providing complementing genes as appropriate, or the appropriate transcription factors for initiation of transcription from a cell status-specific TRE for propagation of the adenovirus. Plasmids are generally introduced into a suitable host cell such as 293 cells using appropriate means of transduction, such as cationic liposomes. Alternatively, *in vitro* ligation of the right and left-hand portions of the adenovirus genome can also be used to construct recombinant adenovirus derivative containing all the replication-essential portions of adenovirus genome. Berkner et al.

(1983) *Nucleic Acid Research* 11: 6003-6020; Bridge et al. (1989) *J. Virol.* 63: 631-638.

For convenience, plasmids are available that provide the necessary portions of adenovirus. Plasmid pXC.1 (McKinnon (1982) *Gene* 19:33-42) contains the wild-type left-hand end of Ad5. pBHG10 (Bett et al. (1994) *Proc. Natl. Acad. Sci USA* 91:8802-8806; Microbix Biosystems Inc., Toronto) provides the right-hand end of Ad5, with a deletion in E3. The deletion in E3 provides room in the virus to insert a 3 kb cell status-TRE without deleting the endogenous enhancer/promoter. Bett et al. (1994). The gene for E3 is located on the opposite strand from E4 (r-strand). pBHG11 provides an even larger E3 deletion (an additional 0.3 kb is deleted). Bett et al. (1994).

For manipulation of the early genes, the transcription start site of Ad5 E1A is at 498 and the ATG start site of the E1A protein is at 560 in the virus genome. This region can be used for insertion of an cell status-specific TRE. A restriction site may be introduced by employing polymerase chain reaction (PCR), where the primer that is employed may be limited to the Ad5 genome, or may involve a portion of the plasmid carrying the Ad5 genomic DNA. For example, where pBR322 is used, the primers may use the EcoRI site in the pBR322 backbone and the XbaI site at 1339 of Ad5. By carrying out the PCR in two steps, where overlapping primers at the center of the region introduce a 30 sequence change resulting in a unique restriction site, one can provide for insertion of heterologous TRE at that site.

A similar strategy may also be used for insertion of a heterologous TRE to regulate E1B. The E1B promoter of Ad5 consists of a single high-affinity recognition site for Spl and a TATA box. This region extends from 1636 to 1701. By insertion of a heterologous TRE in this region, one can provide for target cell-specific transcription of the E1B gene. By employing the left-hand region modified with a heterologous TRE regulating E1A as the template for introducing a heterologous TRE to regulate E1B, the resulting adenovirus vector will be dependent

upon the cell status-specific transcription factors for expression of both E1A and E1B.

Similarly, a cell status-specific TRE can be inserted upstream of the E2 gene to make its expression cell status specific. The E2 early promoter, mapping in Ad5 from 27050-27150, consists of a major and a minor transcription initiation site, the latter accounting for about 5% of the E2 transcripts, two non-canonical TATA boxes, two E2F transcription factor binding sites and an ATF transcription factor binding site. For a detailed review of the E2 promoter architecture see Swaminathan et al., *Curr. Topics in Micro. and Imm.* (1995) 199 (part 3):177-194.

For E4, one must use the right hand portion of the adenovirus genome. The E4 transcription start site is predominantly at 35609, the TATA box at 35638 and the first ATG/CTG of ORF 1 is at 35532. Virtanen et al. (1984) *J. Virol.* 51: 822-831. Using any of the above strategies for the other genes, a cell status-specific TRE may be introduced upstream from the transcription start site. For the construction of mutants in the E4 region, the co-transfection and homologous recombination are performed in W162 cells (Weinberg et al. (1983) *Proc. Natl. Acad. Sci.* 80:5383-5386) which provide E4 proteins *in trans* to complement defects in synthesis of these proteins. Alternatively, these constructs can be produced by *in vitro* ligation.

#### ***Methods using the adenovirus vectors of the invention***

The adenoviral vectors of the invention can be used for a wide variety of purposes, which will vary with the desired or intended result. Accordingly, the present invention includes methods using the adenoviral vectors described above.

In one embodiment, methods are provided for conferring selective cytotoxicity in target cells (i.e., cells exhibiting a requisite physiological state which allows a cell status-specific TRE to function), generally but not necessarily in an individual (preferably human), comprising contacting the cells with an adenovirus vector described herein, such that the adenovirus vector enters the cell. Cytotoxicity can be



measured using standard assays in the art, such as dye exclusion, <sup>3</sup>H-thymidine incorporation, and/or lysis.

In another embodiment, methods are provided for propagating an adenovirus specific for mammalian cells which allow a cell status-specific TRE to function.

5 These methods entail combining an adenovirus vector with mammalian cells, whereby said adenovirus is propagated.

10 The invention further provides methods of suppressing tumor cell growth, generally but not necessarily in an individual (preferably human), comprising contacting a tumor cell with an adenoviral vector of the invention such that the adenoviral vector enters the tumor cell and exhibits selective cytotoxicity for the tumor cell. Tumor cell growth can be assessed by any means known in the art, including, but not limited to, measuring tumor size, determining whether tumor cells are proliferating using a <sup>3</sup>H-thymidine incorporation assay, or counting tumor cells.

15 The invention also includes methods for detecting target cells (i.e., cells which permit or induce a cell status-specific TRE to function) in a biological sample. These methods are particularly useful for monitoring the clinical and/or physiological condition of an individual (i.e., mammal), whether in an experimental or clinical setting. For these methods, cells of a biological sample are contacted with an adenovirus vector, and replication of the adenoviral vector is detected. A suitable  
20 biological sample is one in which cells exhibiting a requisite physiological (and/or environmental) state, for example, an aberrant physiological state (such as cells in hypoxic conditions and exhibiting a phenotype characteristic of cells in hypoxic conditions, such as expression of HIF-1) may be or are suspected to be present. Generally, in mammals, a suitable clinical sample is one in which cancerous cells  
25 exhibiting a requisite physiological state, such as cells within a solid tumor which are under hypoxic conditions, are suspected to be present. Such cells can be obtained, for example, by needle biopsy or other surgical procedure. Cells to be contacted may be treated to promote assay conditions, such as selective enrichment, and/or

solubilization. In these methods, target cells can be detected using *in vitro* assays that detect adenoviral proliferation, which are standard in the art. Examples of such standard assays include, but are not limited to, burst assays (which measure virus yield) and plaque assays (which measure infectious particles per cell). Propagation can also be detected by measuring specific adenoviral DNA replication, which are also standard assays.

The following examples are provided to illustrate but not limit the invention.

## EXAMPLES

### EXAMPLE 1

#### ***Adenovirus vector comprising E1A under transcriptional control of a hypoxia responsive element and a PSA-TRE***

##### *General techniques*

A human embryonic kidney cell line, 293, efficiently expresses E1A and E1B genes of Ad5 and exhibits a high transfection efficiency with adenovirus DNA. To generate recombinant adenovirus, 293 cells were co-transfected with one left end Ad5 plasmid and one right end Ad5 plasmid. Homologous recombination generates adenoviruses with the required genetic elements for replication in 293 cells which provide E1A and E1B proteins *in trans* to complement defects in synthesis of these proteins.

The plasmids to be combined were co-transfected into 293 cells using cationic liposomes such as Lipofectin (DOTMA:DOPE™, Life Technologies) by combining the two plasmids, then mixing the plasmid DNA solution (10 µg of each plasmid in 500 µl of minimum essential medium (MEM) without serum or other additives) with a four-fold molar excess of liposomes in 200 µl of the same buffer. The DNA-lipid complexes were then placed on the cells and incubated at 37°C, 5% CO<sub>2</sub> for 16 hours. After incubation the medium was changed to MEM with 10% fetal bovine serum and

the cells are further incubated at 37°C, 5% CO<sub>2</sub>, for 10 days with two changes of medium. At the end of this time the cells and medium were transferred to tubes, freeze-thawed three times, and the lysate was used to infect 293 cells at the proper dilution to detect individual viruses as plaques.

5           Plaques obtained were plaque purified twice, and viruses were characterized for presence of desired sequences by PCR and occasionally by DNA sequencing. For further experimentation, the viruses were purified on a large scale by cesium chloride gradient centrifugation.

10           *Adenovirus vectors in which E1A is under transcriptional control of a cell status-specific TRE*

          An adenovirus vector containing a hypoxia response element (HRE) was generated. CN796, an adenovirus vector in which E1A is under the control of a composite TRE consisting of an HRE and a PSA-TRE, was made by co-transfecting CN515 with pBHG10. CN515 was constructed by inserting a 67 base pair fragment  
15           from HRE eno1 (Jiang et al. (1997) *Cancer Research* 57:5328-5335) (SEQ ID NO:1; Figure 2) into CN65 at the BglII site. CN65 is a plasmid containing an enhancer and promoter from the human PSA gene, consisting of an enhancer from -5322 to -3738 fused to a PSA promoter from -541 to +12. This is the PSA-TRE contained within plasmid CN706. Rodriguez et al. (1997) *Cancer Res.* 57:2559-2563.

20           *Virus growth in vitro*

          Growth selectivity of recombinant adenovirus is analyzed in plaque assays in which a single infectious particle produces a visible plaque by multiple rounds of infection and replication. Virus stocks are diluted to equal pfu/ml, then used to infect monolayers of cells for 1 hour. The inoculum is then removed and the cells are  
25           overlayed with semisolid agar containing medium and incubated at 37°C for 10 days. Plaques in the monolayer are then counted and titers of infectious virus on the various cells are calculated. The data are normalized to the titer of CN702 (wild type) on 293 cells.

## Claims

What is claimed is:

1. An adenovirus vector comprising an adenovirus gene under transcriptional control of a transcriptional regulatory element (TRE) comprising a cell status-specific TRE.

2. The adenovirus vector of claim 1, wherein the adenovirus gene is essential for viral replication.

3. The adenovirus vector of claim 2, wherein the adenovirus gene is an early gene.

4. The adenovirus vector of claim 2, wherein the adenovirus gene is a late gene.

5. The adenovirus vector of claim 3, wherein the adenovirus early gene is E1A.

6. The adenovirus vector of claim 3, wherein the adenovirus early gene is E1B.

7. The adenovirus vector of claim 3, wherein the adenovirus early gene is E4.

8. The adenovirus vector of claim 1, wherein the cell status-specific TRE is human.

9. The adenovirus vector of claim 1, wherein the cell status-specific TRE comprises a hypoxia responsive element (HRE).

10. The adenovirus vector of claim 9, wherein the HRE comprises SEQ ID NO:1.

11. The adenovirus vector of claim 1, wherein the cell status-specific TRE comprises a cell cycle specific element.

12. The adenovirus vector of claim 11, wherein the cell cycle-specific element is from the E2F-1 gene.

5           13. The adenovirus vector of claim 1, wherein the cell status-specific TRE comprises a heat-inducible element.

14. The adenovirus vector of claim 1, further comprising a cell type-specific TRE.

10           15. The adenovirus vector of claim 14, wherein the cell type-specific TRE is prostate cell specific.

16. The adenovirus vector of claim 15, wherein the prostate cell-specific TRE is a *PSA*-TRE.

17. The adenovirus vector of claim 1, further comprising a transgene under transcriptional control of a second cell status-specific TRE.

15           18. An adenovirus vector comprising an adenovirus gene under transcriptional control of a a TRE comprising a cell status-specific TRE and a cell-type specific TRE.

19 The adenovirus vector of claim 18, wherein the adenovirus gene is an early gene.

20           20. The adenovirus vector of claim 19, wherein the adenovirus early gene is E1A.

21. The adenovirus vector of claim 20, wherein the cell status-specific TRE comprises an HRE and the cell-type specific TRE is a *PSA*-TRE.

22. The adenovirus vector of claim 21, wherein the HRE comprises SEQ ID NO:1 and the PSA-TRE comprises nucleotides about 503 to about 2086 of SEQ ID NO:3 and nucleotides about 5285 to about 5836 of SEQ ID NO:3.

23. A composition comprising an adenovirus vector of claim 1.

5           24. The composition of claim 23, further comprising a pharmaceutically acceptable excipient.

25. A host cell comprising the adenovirus vector of claim 1.

10           26. A method of propagating adenovirus specific for cells which allow a cell status-specific TRE to function, said method comprising combining an adenovirus according to claim 1 with the cells, whereby said adenovirus is propagated.

27. A method for conferring selective cytotoxicity on a target cell, said method comprising contacting a cell which allows a cell status-specific TRE to function with an adenovirus vector of claim 1, whereby the vector enters the cell.

15           28. A method for suppressing tumor growth comprising introducing the adenovirus vector of claim 1 into a tumor cell which allows a cell status-specific TRE to function, wherein introduction of the adenovirus vector results in suppression of tumor growth.

The present invention provides adenoviral vectors comprising cell status-specific transcriptional regulatory elements which confer cell status-specific transcriptional regulation on an adenoviral gene. A “cell status” is generally a reversible physiological and/or environmental state. The invention further provides compositions and host cells comprising the vectors, as well as methods of using the vectors.

□



146

1000

1001

☐

# FIGURE 1

[illegible]



## FIGURE 2

ccccccaggg cagtttgcac ggggttgcac gacttcgacagtcgacacccgc ggggttgcag gcccggac

[illegible]

FIGURE 3

|             |             |             |             |             |             |      |
|-------------|-------------|-------------|-------------|-------------|-------------|------|
| ggggcccaaaa | ttagcaagtg  | accacgtgtg  | tctgaagcca  | gtgacctaa   | gaccaccctt  | 61   |
| gcagaaccgt  | ggtctccttg  | tcacagctca  | ggcagcctct  | ggcttagcct  | ctgttctctt  | 121  |
| cataaccctt  | ctcagcgccct | gctctgggcc  | agaccagtgt  | tgggagagct  | cgctactgag  | 181  |
| ctcctagatt  | ggcagggggag | gcagatggag  | aaaaagagt   | tgtgtgtca   | gcattggagc  | 241  |
| agagggcagca | gtgggcaata  | gaggaagtga  | gtaaatcctt  | gggagggctc  | cctagaagt   | 301  |
| atgtgttttc  | tttttttgtt  | ttagagacag  | gattctcgct  | tgtcgccca   | gctggtgtgc  | 361  |
| agtggcataga | tcatagctca  | ctgcagcctc  | gacttctcgg  | gctcaagcaa  | tcctccacc   | 421  |
| tcagcctccc  | aagtagcttg  | gaactacgggc | acacggccacc | atgcctggct  | aattttgta   | 481  |
| tttttgtag   | agatgggtct  | tcacacatgt  | gatacagctg  | gtctcgaact  | cctgggtcca  | 541  |
| tgcgataccac | cccgccagct  | gattacaggg  | atccggttg   | tgaugccacc  | cgcccaagac  | 601  |
| ccacttcac   | gtattgtaaa  | cgctctgtac  | cttctgttc   | ccctgtctac  | tggactgtga  | 661  |
| gctccttaag  | gccacgaatt  | gaggaatggg  | cacagaagcaa | gctctccaaa  | cgttgttga   | 721  |
| atgagtgaag  | gaatgaatga  | gttcaagcag  | atgctatacg  | ttggctgttg  | gagatlttg   | 781  |
| ctaaaaatggg | acttgcaaga  | aaagcccgac  | tccccctcgc  | catttccagg  | caccgctctt  | 841  |
| cagcttgggc  | tctgggtgag  | cgggataagg  | ctgggtgcag  | gattagata   | atgtcatggg  | 901  |
| tgaaggcaagt | tgaggaatga  | agaagtggct  | gatggctggg  | ctgtggaact  | gatgatcctg  | 961  |
| aaaagaaagag | gggacagctc  | ctggaaatct  | aaagctgaag  | tgttggggc   | tacaggttga  | 1021 |
| gggtcacgtg  | cagaagaaag  | gctctgttct  | gaaacctgcac | tatagaagg   | tcaagtggat  | 1081 |
| ggcggagucgt | cggggcgggg  | cggggacctat | gttcccgtgt  | ccccacgcct  | ccaagcaggg  | 1141 |
| acgccccgggc | tgggggcggg  | gagtcagacc  | ggcgcctgta  | ccatccggac  | aaagcctgcg  | 1201 |
| cgcgccccgcg | ccccccattg  | ggcgtaacgc  | ccccggccgc  | cgccccatcc  | cgccccctgc  | 1261 |
| cgccgggtccc | ggcgcgttaa  | agccaatagg  | aaccggccgc  | gttgttccg   | tcacggccgg  | 1321 |
| ggcagcccaat | tgtggcggcg  | ctcgcggtct  | cggtgtctct  | tcgcyggcaaa | aaagatlttg  | 1381 |
| cgcgtaaaag  | *tggcggggac | tttgcaggca  | ggcggcggcc  | ggggcggagc  | gggatcgaagc | 1441 |
| cctcgcccgag | gcctgcggcc  | atggggccgc  | ggcggcggcc  | ccgectgtca  | cccgggccgc  | 1501 |
| ggcgggccgtg | agcgtcatatg |             |             |             |             |      |

[illegible]

|            |            |            |            |             |             |      |
|------------|------------|------------|------------|-------------|-------------|------|
| aagcttctag | ttttcttttc | ccggtgacat | cgtggaaagc | actagcatct  | ctaagcaatg  | 60   |
| atctgtgaca | atattcacag | tgtaatgcc  | tccagggaac | tcaactgagc  | ottgatgtcc  | 120  |
| agagatTTTT | gtgttttttt | ctgagactga | gtctcgctct | gtgccaggct  | ggagtgcagt  | 180  |
| ggtgcaacct | tggctcactg | caagctccgc | ctcctggggt | cacgccattc  | tcttgccctca | 240  |
| gcctcctgag | tagctgggac | tacaggcacc | cgccaccacg | cctggctaata | ttttttgtat  | 300  |
| ttttagtaga | gatgggggtt | cactgtgtta | gccaggatgg | tctcagtctc  | ctgacctcgt  | 360  |
| gatctgcccc | ccttggcctc | ccaaagtgct | gggatgacag | gcgtgagcca  | cgcgcctgg   | 420  |
| ccgatatcca | gagatTTTT  | ggggggctcc | atcacacaga | catgttgact  | gtcttcatgg  | 480  |
| ttgactttta | gtatccagcc | cctctagaaa | tctagctgat | atagtgtggc  | tcaaaaacct  | 540  |
| cagcaciaat | cacaccgtta | gactatctgg | tgtggcccaa | accttcagggt | gaacaaaggg  | 600  |
| actctaactc | ggcaggatat | tccaaagcat | tagagatgac | ctcttgcaaa  | gaaaaagaaa  | 660  |
| tggaaaagaa | aaagaaagaa | aggaaaaaaa | aaaaaaaaaa | gagatgacct  | ctcaggctct  | 720  |
| gaggggaaac | gcctgaggtc | tttgagcaag | gtcagtcctc | tgttgccacag | tctccctcac  | 780  |
| agggtcattg | tgacgatcaa | atgtggtcac | gtgtatgagg | caccagcaca  | tgcttggtct  | 840  |
| tggggagtgc | cgtgtaagtg | tatgcttgca | ctgctgaatg | cttgggatgt  | gtcagggatt  | 900  |
| atcttcagca | cttacagatg | ctcatctcat | cctcacagca | tactatggg   | atgggtatta  | 960  |
| ctggcctcat | ttgatggaga | aagtggctgt | ggctcagaaa | ggggggacca  | ctagaccagg  | 1020 |
| gacactctgg | atgctgggga | ctccagagac | catgaccact | caccaactgc  | agagaaatta  | 1080 |
| attgtggcct | gatgtccctg | tcctggagag | ggaggagggt | gaccttact   | aacctctac   | 1140 |
| cttgaccctc | tcttttaggg | ctctttctga | cctccaccat | ggtactagga  | ccccattgta  | 1200 |
| ttctgtacct | tcttgactct | atgaccccc  | ctgcccactg | catccagctg  | ggtccctcc   | 1260 |
| tatctctatt | cccagctggc | cagtgcagtc | tcagtgccca | cctgtttgtc  | agtaactctg  | 1320 |
| aaggggctga | cattttactg | acttgcaaac | aaataagcta | actttccaga  | gttttgtgaa  | 1380 |
| tgctggcaga | gtccatgaga | ctcctgagtc | agaggcaaag | gcttttactg  | ctcacagctt  | 1440 |
| agcagacagc | atgaggttca | tgttcacatt | agtacacctt | gcccccccca  | aatctttag   | 1500 |
| ggtgaccaga | gcagtctagg | tggatgctgt | gcagaagggg | tttgtgccac  | tggtgagaaa  | 1560 |
| cctgagatta | ggaatcctca | atcttatact | gggacaactt | gcaaacctgc  | tcagcctttg  | 1620 |
| tctctgatga | agatattatc | ttcatgatct | tggattgaaa | acagacctac  | tctggaggaa  | 1680 |

## FIGURE 4B

catattgtat cgattgtcct tgacagtaaa caaatctggt gtaagagaca ttatctttat 1740  
 tatctaggac agtaagcaag cctggatctg agagagatat catcttgcaa ggatgcctgc 1800  
 ttacaaaaca tccttgaaac aacaatccag aaaaaaaaaag gtgttgctgt ctttgctcag 1860  
 aagacacaca gatacgtgac agaaccatgg agaattgcct cccaacgctg ttcagccaga 1920  
 gccttccacc cttgtctgca ggacagtctc aacgttccac cattaaatac ttcttctatc 1980  
 acatcctgct tctttatgcc taaccaaggt tctaggtccc gatcgactgt gtctggcagc 2040  
 actccactgc caaaccaga ataaggcagc gctcaggatc ccgaaggggc atggctgggg 2100  
 atcagaactt ctgggttga gtgaggagt ggtccaccct cttgaatttc aaaggaggaa 2160  
 gaggctggat gtgaaggtag tgggggaggg aaagtgtcag ttccgaactc ttaggtcaat 2220  
 gagggaggag actggttaagg tcccagctcc cgaggtagct atgtgggaat ggcctaagaa 2280  
 tctcatatcc tcaggaagaa ggtgctggaa tcctgagggg tagagtctct ggtatatttg 2340  
 tggcttaagg ctctttggcc cctgaaggca gaggctggaa ccattaggtc cagggtttgg 2400  
 ggtgatagta atgggatctc ttgattcctc aagagtctga ggatcgaggg ttgcccattc 2460  
 ttccatcttg ccacctaata cttactccac ttgagggtag caccagccct tctagctcca 2520  
 tgaagggtccc ctgggcaagc acaatctgag catgaaagat gcccagagg ccttgggtgt 2580  
 catccactca tcatccagca tcacactctg aggggtgtggc cagcaccatg acgtcatgtt 2640  
 gctgtgacta tccctgcagc gtgcctctcc agccacctgc caaccgtaga gctgcccata 2700  
 ctctcttggt gggagtggcc tgcattgtgc caggctgagg cctagtgtca gacagggagc 2760  
 ctggaatcat agggatccag gactcaaaaag tgctagagaa tggccatatg tcaccatcca 2820  
 tgaaatctca agggcttctg ggtggagggc acagggacct gaacttatgg tttcccaagt 2880  
 ctattgctct cccaagttag tctcccagat acgaggcact gtgccagcat cagccttata 2940  
 tccaccacat cttgtaaaag gactaccag ggccctgatg aacaccatgg tgtgtacagg 3000  
 agtagggggg ggaggcacgg actcctgtga ggtcacagcc aaggagcat catcatgggt 3060  
 ggggaggagg caatggacag gcttgagaac ggggatgtgg ttgtatttg tttcttttg 3120  
 ttagataaag tgctgggtat aggattgaga gtggagtatg aagaccagtt aggatggagg 3180  
 atcagattgg agttgggtta gataaagtgc tgggtatagg attgagagtg gagtatgaag 3240  
 accagttagg atggaggatc agattggagt tgggttagag atggggtaaa attgtgctcc 3300  
 ggatgagttt gggattgaca ctgtggaggt gggttgggat ggcattggct tgggatggaa 3360

# FIGURE 4C

atagatttgt tttgatgttg gctcagacat ccttggggat tgaactgggg atgaagctgg 3420  
 gtttgatttt ggaggtagaa gacgtggaag tagctgtcag atttgacagt ggccatgagt 3480  
 tttgtttgat ggggaatcaa acaatggggg aagacataag ggttggcttg ttaggttaag 3540  
 ttgcgttggg ttgatggggg cggggctgtg tataatgcag ttggattggg ttgtattaaa 3600  
 ttgggttggg tcaggttttg gttgaggatg agttgaggat atgcttgggg acaccggatc 3660  
 catgaggttc tcaactggagt ggagacaaac ttcctttcca ggatgaatcc agggaagcct 3720  
 taattcacgt gtaggggagg tcaggccact ggctaagtat atccttccac tccagctcta 3780  
 agatggtcct aaattgtgat tatctatata cacttctgtc tccctcaactg tgcttggagt 3840  
 ttacctgatc actcaactag aaacagggga agattttatc aaattctttt tttttttttt 3900  
 ttttttttga gacagagtct cactctgttg cccaggctgg agtgcagtgg cgcagtctcg 3960  
 gctcaactga acctctgcct cccaggttca agtgattctc ctgcctcagc ctcttgagtt 4020  
 gctgggatta caggcatgca gcaccatgcc cagctaattt ttgtattttt agtagagatg 4080  
 gggtttcacc aatgtttgcc aggtggcct cgaactcctg acctggtgat ccacctgcct 4140  
 cagcctccca aagtgtctggg attacaggcg tcagccaccg cgcccagcca cttttgtcaa 4200  
 attcttgaga cacagctcgg gctggatcaa gtgagctact ctggttttat tgaacagctg 4260  
 aaataaccaa ctttttgga attgatgaaa tcttacggag ttaacagtgg aggtaccagg 4320  
 gctcttaaga gttcccgatt ctcttctgag actacaaatt gtgattttgc atgccacctt 4380  
 aatctttttt tttttttttt taaatcgagg tttcagtctc attctatttc ccaggctgga 4440  
 gttcaatagc gtgatcacag ctcaactgtag ccttgaactc ctggccttaa gagattctcc 4500  
 tgcttcgggc tcccaatagc taagactaca gtagtccacc accatatcca gataattttt 4560  
 aaattttttg gggggccggg cacagtggct cagcctgta atcccaacac catgggaggc 4620  
 tgagatgggt ggatcacgag gtcaggagtt tgagaccagc ctgaccaaca tggtgaaact 4680  
 ctgtctctac taaaaaaaaa aaaaatagaa aaattagccg ggcgtggtgg cacacggcac 4740  
 ctgtaatccc agctactgag gaggctgagg caggagaatc acttgaaccc agaaggcaga 4800  
 ggttgcaatg agccgagatt gcgccactgc actccagcct gggtgacaga gtgagactct 4860  
 gtctcaaaaa aaaaaaattt tttttttttt tttgtagaga tggatcttgc tttgtttctc 4920  
 tggttggcct tgaactcctg gcttcaagtg atcctcctac cttggcctcg gaaagtgttg 4980  
 ggattacagg cgtgagccac catgactgac ctgtcgttaa tcttgaggta cataaacctg 5040  
 gctcctaaag gctaaaggct aaatatttgt tggagaaggg gcattggatt ttgcatgagg 5100

3420  
 3480  
 3540  
 3600  
 3660  
 3720  
 3780  
 3840  
 3900  
 3960  
 4020  
 4080  
 4140  
 4200  
 4260  
 4320  
 4380  
 4440  
 4500  
 4560  
 4620  
 4680  
 4740  
 4800  
 4860  
 4920  
 4980  
 5040  
 5100

## FIGURE 4D

atgattctga cctgggaggg caggtcagca ggcatctctg ttgcacagat agagtgtaca 5160  
ggtctggaga acaaggagtg gggggttatt ggaattccac attgtttgct gcacgttgga 5220  
ttttgaaatg ctagggaact ttgggagact catatttctg ggctagagga tctgtggacc 5280  
acaagatcct tttatgatga cagtagcaat gtatctgtgg agctggattc tgggttggga 5340  
gtgcaaggaa aagaatgtac taaatgcaa gacatctatt tcaggagcat gaggaataaa 5400  
agttctagtt tctggtctca gagtgggtga gggatcaggg agtctcaca tctcctgagt 5460  
gctggtgtct tagggcacac tgggtcttgg agtgcaaagg atctaggcac gtgaggcttt 5520  
gtatgaagaa tcggggatcg taccacccc ctgtttctgt ttcattctgg gcatgtctcc 5580  
tctgcctttg tcccctagat gaagtctcca tgagctacaa ggcctggtg catccagggt 5640  
gatctagtaa ttgcagaaca gcaagtgcta gctctccctc cccttcaca gctctgggtg 5700  
tgggaggggg ttgtccagcc tccagcagca tggggagggc cttggtcagc ctctgggtgc 5760  
cagcagggca ggggcggagt cctggggaat gaaggtttta tagggctcct gggggaggct 5820  
ccccagcccc aagctt 5836

bioRxiv preprint doi: <https://doi.org/10.1101/2021.03.10.433333>; this version posted March 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

|  | 1990 | 1991 | 1992 | 1993 | 1994 | 1995 | 1996 | 1997 | 1998 | 1999 | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 | 2023 | 2024 | 2025 | 2026 | 2027 | 2028 | 2029 | 2030 | 2031 | 2032 | 2033 | 2034 | 2035 | 2036 | 2037 | 2038 | 2039 | 2040 | 2041 | 2042 | 2043 | 2044 | 2045 | 2046 | 2047 | 2048 | 2049 | 2050 | 2051 | 2052 | 2053 | 2054 | 2055 | 2056 | 2057 | 2058 | 2059 | 2060 | 2061 | 2062 | 2063 | 2064 | 2065 | 2066 | 2067 | 2068 | 2069 | 2070 | 2071 | 2072 | 2073 | 2074 | 2075 | 2076 | 2077 | 2078 | 2079 | 2080 | 2081 | 2082 | 2083 | 2084 | 2085 | 2086 | 2087 | 2088 | 2089 | 2090 | 2091 | 2092 | 2093 | 2094 | 2095 | 2096 | 2097 | 2098 | 2099 | 2100 | 2101 | 2102 | 2103 | 2104 | 2105 | 2106 | 2107 | 2108 | 2109 | 2110 | 2111 | 2112 | 2113 | 2114 | 2115 | 2116 | 2117 | 2118 | 2119 | 2120 | 2121 | 2122 | 2123 | 2124 | 2125 | 2126 | 2127 | 2128 | 2129 | 2130 | 2131 | 2132 | 2133 | 2134 | 2135 | 2136 | 2137 | 2138 | 2139 | 2140 | 2141 | 2142 | 2143 | 2144 | 2145 | 2146 | 2147 | 2148 | 2149 | 2150 | 2151 | 2152 | 2153 | 2154 | 2155 | 2156 | 2157 | 2158 | 2159 | 2160 | 2161 | 2162 | 2163 | 2164 | 2165 | 2166 | 2167 | 2168 | 2169 | 2170 | 2171 | 2172 | 2173 | 2174 | 2175 | 2176 | 2177 | 2178 | 2179 | 2180 | 2181 | 2182 | 2183 | 2184 | 2185 | 2186 | 2187 | 2188 | 2189 | 2190 | 2191 | 2192 | 2193 | 2194 | 2195 | 2196 | 2197 | 2198 | 2199 | 2200 | 2201 | 2202 | 2203 | 2204 | 2205 | 2206 | 2207 | 2208 | 2209 | 2210 | 2211 | 2212 | 2213 | 2214 | 2215 | 2216 | 2217 | 2218 | 2219 | 2220 | 2221 | 2222 | 2223 | 2224 | 2225 | 2226 | 2227 | 2228 | 2229 | 2230 | 2231 | 2232 | 2233 | 2234 | 2235 | 2236 | 2237 | 2238 | 2239 | 2240 | 2241 | 2242 | 2243 | 2244 | 2245 | 2246 | 2247 | 2248 | 2249 | 2250 | 2251 | 2252 | 2253 | 2254 | 2255 | 2256 | 2257 | 2258 | 2259 | 2260 | 2261 | 2262 | 2263 | 2264 | 2265 | 2266 | 2267 | 2268 | 2269 | 2270 | 2271 | 2272 | 2273 | 2274 | 2275 | 2276 | 2277 | 2278 | 2279 | 2280 | 2281 | 2282 | 2283 | 2284 | 2285 | 2286 | 2287 | 2288 | 2289 | 2290 | 2291 | 2292 | 2293 | 2294 | 2295 | 2296 | 2297 | 2298 | 2299 | 2300 | 2301 | 2302 | 2303 | 2304 | 2305 | 2306 | 2307 | 2308 | 2309 | 2310 | 2311 | 2312 | 2313 | 2314 | 2315 | 2316 | 2317 | 2318 | 2319 | 2320 | 2321 | 2322 | 2323 | 2324 | 2325 | 2326 | 2327 | 2328 | 2329 | 2330 | 2331 | 2332 | 2333 | 2334 | 2335 | 2336 | 2337 | 2338 | 2339 | 2340 | 2341 | 2342 | 2343 | 2344 | 2345 | 2346 | 2347 | 2348 | 2349 | 2350 | 2351 | 2352 | 2353 | 2354 | 2355 | 2356 | 2357 | 2358 | 2359 | 2360 | 2361 | 2362 | 2363 | 2364 | 2365 | 2366 | 2367 | 2368 | 2369 | 2370 | 2371 | 2372 | 2373 | 2374 | 2375 | 2376 | 2377 | 2378 | 2379 | 2380 | 2381 | 2382 | 2383 | 2384 | 2385 | 2386 | 2387 | 2388 | 2389 | 2390 | 2391 | 2392 | 2393 | 2394 | 2395 | 2396 | 2397 | 2398 | 2399 | 2400 | 2401 | 2402 | 2403 | 2404 | 2405 | 2406 | 2407 | 2408 | 2409 | 2410 | 2411 | 2412 | 2413 | 2414 | 2415 | 2416 | 2417 | 2418 | 2419 | 2420 | 2421 | 2422 | 2423 | 2424 | 2425 | 2426 | 2427 | 2428 | 2429 | 2430 | 2431 | 2432 | 2433 | 2434 | 2435 | 2436 | 2437 | 2438 | 2439 | 2440 | 2441 | 2442 | 2 |
|--|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|---|
|--|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|---|

|            |             |             |            |             |             |      |
|------------|-------------|-------------|------------|-------------|-------------|------|
| aagcttttta | gtgctttaga  | cagtgagctg  | gtctgtctaa | cccaagtgac  | ctgggctcca  | 60   |
| tactcagccc | cagaagtgaa  | gggtgaagct  | gggtggagcc | aaaccaggca  | agcctaccct  | 120  |
| cagggctccc | agtggcctga  | gaaccattgg  | acccaggacc | cattacttct  | agggtaagga  | 180  |
| aggtacaaac | accagatcca  | accatggtct  | ggggggacag | ctgtcaaattg | cctaaaaata  | 240  |
| tacctgggag | aggagcaggc  | aaactatcac  | tgccccaggt | tctctgaaca  | gaaacagagg  | 300  |
| ggcaacccaa | agtccaaatc  | caggtgagca  | ggtgcaccaa | atgccagag   | atatgacgag  | 360  |
| gcaagaagtg | aaggaaccac  | ccctgcatca  | aatgttttgc | atgggaagga  | gaaggggggtt | 420  |
| gctcatgttc | ccaatccagg  | agaatgcatt  | tgggatctgc | cttctttctca | ctccttggtt  | 480  |
| agcaagacta | agcaaccagg  | actctggatt  | tggggaaaga | cgtttatattg | tggagggccag | 540  |
| tgatgacaat | cccacgaggg  | cctaggtgaa  | gagggcagga | aggctcgaga  | cactggggac  | 600  |
| tgagtgaaaa | ccacacccat  | gatctgcacc  | acccatggat | gtccttcat   | tgtcacctt   | 660  |
| tctgttgata | tcagatggcc  | ccattttctg  | taccttcaca | gaaggacaca  | ggctagggtc  | 720  |
| tgtgcatggc | cttcatcccc  | ggggccatgt  | gaggacagca | ggtgggaaag  | atcatgggtc  | 780  |
| ctcctgggtc | ctgcagggcc  | agaacattca  | tcaccatac  | tgacctccta  | gatgggaatg  | 840  |
| gcttccctgg | ggctggggcca | acggggcctg  | ggcaggggag | aaaggacgtc  | aggggacagg  | 900  |
| gaggaaggg  | catcgagacc  | cagcctggaa  | ggttcttgtc | tctgaccatc  | caggatttac  | 960  |
| ttccctgcat | ctacctttgg  | tcattttccc  | tcagcaatga | ccagctctgc  | ttcctgatct  | 1020 |
| cagcctccca | ccctggacac  | agcaccaccag | tccctggccc | ggctgcatcc  | acccaatacc  | 1080 |
| ctgataaccc | aggacccatt  | acttctaggg  | taaggagggt | ccaggagaca  | gaagctgagg  | 1140 |
| aaaggtctga | agaagtcaca  | tctgtcctgg  | ccagagggga | aaaaccatca  | gatgctgaac  | 1200 |
| caggagaatg | ttgaccacag  | aaagggaccg  | aggacccaag | aaaggagtca  | gaccaccagg  | 1260 |
| gtttgcctga | gaggaaggat  | caaggccccg  | agggaaagca | gggctggctg  | catgtgcagg  | 1320 |
| acactggtgg | ggcatatgtg  | tcttagattc  | tccctgaatt | cagtgtccct  | gccatggcca  | 1380 |
| gactctctac | tcaggcctgg  | acatgctgaa  | ataggacaat | ggccttgtcc  | tctctcccca  | 1440 |
| ccatttgga  | agagacataa  | aggacattcc  | aggacatgcc | ttcctgggag  | gtccagggttc | 1500 |
| tctgtctcac | acctcaggga  | ctgtagttag  | tgcacagcc  | atggtaggtg  | ctgatctcac  | 1560 |
| ccagcctgtc | caggcccttc  | cactctccac  | tttgtgacca | tgtccaggac  | caccctcag   | 1620 |
| atcctgagcc | tgcaaaatacc | cccttgctgg  | gtgggtggat | tcagtaaaca  | gtgaqctcct  | 1680 |

## FIGURE 5B

|            |             |            |            |             |             |      |
|------------|-------------|------------|------------|-------------|-------------|------|
| atccagcccc | cagagccacc  | tctgtcacct | tcctgctggg | catcatccca  | ccttcacaag  | 1740 |
| cactaaagag | catggggaga  | cctggctagc | tgggtttctg | catcaciaag  | aaaataatcc  | 1800 |
| cccaggttcg | gattcccagg  | gctctgtatg | tggagctgac | agacctgagg  | ccaggagata  | 1860 |
| gcagaggtca | gccctaggga  | gggtgggtca | tccaccagg  | ggacaggggt  | gcaccagcct  | 1920 |
| tgctactgaa | agggcctccc  | caggacagcg | ccatcagccc | tgccctgagag | ctttgctaaa  | 1980 |
| cagcagtcag | aggaggccat  | ggcagtggct | gagctcctgc | tccaggcccc  | aacagaccag  | 2040 |
| accaacagca | caatgcagtc  | cttccccaac | gtcacaggtc | accaaaggga  | aactgagggtg | 2100 |
| ctacctaac  | ttagagccat  | caggggagat | aacagcccaa | tttcccaaac  | aggccagttt  | 2160 |
| caatcccatg | acaatgacct  | ctctgctctc | attcttccca | aaataggacg  | ctgattctcc  | 2220 |
| cccaccatgg | atttctccct  | tgtcccggga | gccttttctg | ccccctatga  | tctgggcact  | 2280 |
| cctgacacac | acctcctctc  | tggtgacata | tcagggtccc | tactgtcaa   | gcagtccaga  | 2340 |
| aaggacagaa | ccttggacag  | cgcccatctc | agcttcaccc | ttctctcttc  | acagggttca  | 2400 |
| gggcaaagaa | taaatggcag  | aggccagtga | gccagagat  | ggtgacaggc  | agtgaccag   | 2460 |
| gggcagatgc | ctggagcagg  | agctggcggg | gccacaggga | gaaggatgatg | caggaaggga  | 2520 |
| aaccagaaa  | tgggcaggaa  | aggaggacac | aggctctgtg | gggctgcagc  | ccagggttgg  | 2580 |
| actatgagtg | tgaagccatc  | tcagcaagta | aggccaggtc | ccatgaacaa  | gagtgggagc  | 2640 |
| acgtggcttc | ctgctctgta  | tatggggtgg | gggattccat | gccccataga  | accagatggc  | 2700 |
| cggggttcag | atggagaagg  | agcaggacag | gggatcccca | ggataggagg  | accccagtgt  | 2760 |
| ccccaccag  | gcagggtgact | gatgaatggg | catgcagggt | cctcctgggc  | tgggctctcc  | 2820 |
| ctttgtccct | caggattcct  | tgaaggaaca | tccggaagcc | gaccacatct  | acctgggtggg | 2880 |
| ttctggggag | tccatgtaaa  | gccaggagct | tgtgttgcta | ggaggggtca  | tggcatgtgc  | 2940 |
| tgggggcacc | aaagagagaa  | acctgagggc | aggcaggacc | tggctctgagg | aggcatggga  | 3000 |
| gccagatgg  | ggagatggat  | gtcaggaaag | gctgccccat | cagggagggt  | gatagcaatg  | 3060 |
| gggggtctgt | gggagtgggc  | acgtgggatt | ccttgggctc | tgccaagttc  | cctcccatag  | 3120 |
| tcacaacctg | gggacactgc  | ccatgaaggg | gcgcctttgc | ccagccagat  | gctgctggtt  | 3180 |
| ctgccccatc | actacctct   | ctgctccagc | cactctgggt | ctttctccag  | atgccctgga  | 3240 |
| cagccctggc | ctgggcctgt  | cccctgagag | gtgttgggag | aagctgagtc  | tctggggaca  | 3300 |
| ctctcatcag | agtctgaaag  | gcacatcagg | aaacatccct | ggtctccagg  | actaggcaat  | 3360 |



# FIGURE 5C

|            |             |             |             |             |            |      |
|------------|-------------|-------------|-------------|-------------|------------|------|
| gaggaaaggg | ccccagctcc  | tccctttgcc  | actgagaggg  | tcgaccctgg  | gtggccacag | 3420 |
| tgacttctgc | gtctgtccca  | gtcaccctga  | aaccacaaca  | aaaccccagc  | cccagaccct | 3480 |
| gcaggtacaa | tacatgtggg  | gacagtctgt  | accaggggga  | agccagttct  | ctcttcctag | 3540 |
| gagaccgggc | ctcagggctg  | tgcccggggc  | aggcgggggc  | agcacgtgcc  | tgtccttgag | 3600 |
| aactcgggac | cttaagggtc  | tctgctctgt  | gaggcacagc  | aaggatcctt  | ctgtccagag | 3660 |
| atgaaagcag | ctcctgcccc  | tcctctgacc  | tcttctctct  | tcccaaactct | caaccaacaa | 3720 |
| ataggtgttt | caaattctcat | catcaaactct | tcatccatcc  | acatgagaaa  | gcttaaaacc | 3780 |
| caatggattg | acaacatcaa  | gagttggaac  | aagtggacat  | ggagatgtta  | cttgtggaaa | 3840 |
| tttagatgtg | ttcagctatc  | gggcaggaga  | atctgtgtca  | aattccagca  | tggttcagaa | 3900 |
| gaatcaaaaa | gtgtcacagt  | ccaaatgtgc  | aacagtgcag  | gggataaaac  | tgtggtgcat | 3960 |
| tcaaactgag | ggatattttg  | gaacatgaga  | aaggaaggga  | ttgctgctgc  | acagaacatg | 4020 |
| gatgatctca | cacatagagt  | tgaaagaaa   | gagtcaatcg  | cagaatagaa  | aatgatcact | 4080 |
| aattccacct | ctataaagtt  | tccaagagga  | aaacccaatt  | ctgctgctag  | agatcagaat | 4140 |
| ggaggtgacc | tgtgccttgc  | aatggctgtg  | agggtcacgg  | gagtgtcact  | tagtgcaggc | 4200 |
| aatgtgccgt | atcttaactct | gggcagggct  | ttcatgagca  | cataggaatg  | cagacattac | 4260 |
| tgctgtgttc | attttacttc  | accggaaaag  | aagaataaaa  | tcagccgggc  | gcggtggctc | 4320 |
| acgcctgtaa | tcccagcact  | ttagaaggct  | gaggtgggca  | gattacttga  | ggtcaggagt | 4380 |
| tcaagaccac | cctggccaat  | atggtgaaac  | cccggctcta  | ctaaaaatac  | aaaaattagc | 4440 |
| tgggcatggt | ggtgcgcgcc  | tgtaatccca  | gctactcggg  | aggetgaggc  | tggacaattg | 4500 |
| cttggaacca | ggaagcagag  | gttgacgtga  | gccaaagattg | tgccactgca  | ctccagcttg | 4560 |
| ggcaacagag | ccagactctg  | taaaaaaaaa  | aaaaaaaaaa  | aaaaaaagaa  | agaaagaaaa | 4620 |
| agaaaagaaa | gtataaaatc  | tctttggggt  | aacaaaaaaa  | gatccacaaa  | acaaacacca | 4680 |
| gctcttatca | aacttacaca  | actctgccag  | agaacaggaa  | acacaaatac  | tcattaactc | 4740 |
| acttttgtgg | caataaaacc  | ttcatgtcaa  | aaggagacca  | ggacacaatg  | aggaagtaaa | 4800 |
| actgcaggcc | ctacttgggt  | gcagagaggg  | aaaatccaca  | aataaaacat  | taccagaagg | 4860 |
| agctaagatt | tactgcattg  | agttcattcc  | ccaggtatgc  | aaggtgattt  | taacacctga | 4920 |
| aatcaatca  | ttgcctttac  | tacatagaca  | gattagctag  | aaaaaaatta  | caactagcag | 4980 |
| aacagaagca | atttggcctt  | cctaaaaattc | cacatcatat  | catcatgatg  | gagacagtgc | 5040 |
| agacgccaat | gacaataaaa  | agagggacct  | ccgtcacccg  | gtaaacaatgt | ccacacagct | 5100 |

## FIGURE 5D

|             |            |             |            |             |            |      |
|-------------|------------|-------------|------------|-------------|------------|------|
| ccagcaagca  | cccgtcttcc | cagtgaatca  | ctgtaacctc | ccctttaatc  | agccccaggc | 5160 |
| aaggctgcct  | gcgatggcca | cacaggctcc  | aaccogtggg | cctcaacctc  | ccgcagaggc | 5220 |
| tctcctttgg  | ccaccccatg | gggagagcat  | gaggacaggg | cagagccctc  | tgatgccac  | 5280 |
| acatggcagg  | agctgacgcc | agagccatgg  | gggctggaga | gcagagctgc  | tggggtcaga | 5340 |
| gcttcctgag  | gacaccaggg | cctaagggaa  | ggcagctccc | tggatggggg  | caaccaggct | 5400 |
| ccgggctcca  | acctcagagc | ccgcatggga  | ggagccagca | ctctaggcct  | ttcctagggt | 5460 |
| gactctgagg  | ggaccctgac | acgacaggat  | cgctgaatgc | acccgagatg  | aaggggccac | 5520 |
| cacgggaccc  | tgctctcgtg | gcagatcagg  | agagagtggg | acaccatgcc  | aggcccccat | 5580 |
| ggcatggctg  | cgactgaccc | aggccactcc  | cctgcatgca | tcagcctcgg  | taagtcacat | 5640 |
| gaccaagccc  | aggaccaatg | tggaagggaag | gaaacagcat | cccctttagt  | gatggaaccc | 5700 |
| aaggtcagt   | caaagagagg | ccatgagcag  | ttaggaaggg | tggccaacc   | tacagcacia | 5760 |
| accatcatct  | atcataagta | gaagccctgc  | tccatgaccc | ctgcatttaa  | ataaacgttt | 5820 |
| gttaaagt    | tcaaattccc | tcacatgag   | agctcacctg | tgtgtaggcc  | catcacacac | 5880 |
| acaaacacac  | acacacacac | acacacacac  | acacacacac | acagggaaaag | tgaggatcc  | 5940 |
| tggacagcac  | caggcaggct | tcacaggcag  | agcaaacagc | gtgaatgacc  | catgcagtgc | 6000 |
| cctgggcccc  | atcagctcag | agaccctgtg  | agggtgaga  | tggggctagg  | caggggagag | 6060 |
| acttagagag  | ggtggggcct | ccagggaggg  | ggctgcaggg | agctgggtac  | tgccctccag | 6120 |
| ggagggggct  | gcagggagct | gggtactgcc  | ctccagggag | ggggctgcag  | ggagctgggt | 6180 |
| actgccctcc  | agggaggggg | ctgcagggag  | ctgggtactg | ccctccaggg  | agggggctgc | 6240 |
| agggagctgg  | gtactgccct | ccagggaggc  | aggagcactg | ttccaacag   | agagcacatc | 6300 |
| ttcctgcagc  | agctgcacag | acacaggagc  | ccccatgact | gccctgggcc  | aggggtgtga | 6360 |
| ttccaaattt  | cgtgccccat | tgggtgggac  | ggaggttgac | cgtgacatcc  | aaggggcac  | 6420 |
| tgtgattcca  | aacttaaact | actgtgccta  | caaaatagga | aataacccta  | ctttttctac | 6480 |
| tatctcaaat  | tccctaagca | caagctagca  | ccctttaaat | caggaagttc  | agtcactcct | 6540 |
| ggggctcctcc | catgccccca | gtctgacttg  | caggtgcaca | gggtggctga  | catctgtcct | 6600 |
| tgctcctcct  | cttggctcaa | ctgccgcccc  | tcctgggggt | gactgatggg  | caggacaagg | 6660 |
| gatcctagag  | ctggccccat | gattgacagg  | aaggcaggac | ttggcctcca  | ttctgaagac | 6720 |
| taggggtgtc  | aagagagctg | ggcatccccc  | agagctgcac | aagatgacgc  | ggacagaggg | 6780 |

# FIGURE 5E

|            |            |            |             |             |            |      |
|------------|------------|------------|-------------|-------------|------------|------|
| tgacacaggg | ctcagggctt | cagacgggtc | gggagggtca  | gctgagagtt  | cagggacaga | 6840 |
| cctgaggagc | ctcagtggga | aaagaagcac | tgaagtggga  | agttctggaa  | tgttctggac | 6900 |
| aagcctgagt | gctctaagga | aatgctccca | ccccgatgta  | gcctgcagca  | ctggacggtc | 6960 |
| tgtgtacctc | cccgtgccc  | atcctctcac | agccccgcc   | tctagggaca  | caactcctgc | 7020 |
| cctaacatgc | atctttcctg | tctcattcca | cacaaaagg   | cctctggggt  | ccctgttctg | 7080 |
| cattgcaagg | agtggaggtc | acgttccccc | agaccaccca  | gcaacagggt  | cctatggagg | 7140 |
| tgcggtcagg | aggatcacac | gtccccccat | gccaggggga  | ctgactctgg  | gggtgatgga | 7200 |
| ttggcctgga | ggccactggg | cccctctgtc | cctgagggga  | atctgcaccc  | tggaggctgc | 7260 |
| cacatccctc | ctgattcttt | cagctgaggg | cccttcttga  | aatcccagg   | aggactcaac | 7320 |
| cccactggg  | aaaggcccag | tgtggacggg | tccacagcag  | cccagctaag  | gcccttggac | 7380 |
| acagatcctg | agtgagagaa | cctttaggga | cacagggtga  | cggccatgtc  | cccagtgtcc | 7440 |
| acacagagca | ggggcatctg | gaccctgagt | gtgtagctcc  | cgcgactgaa  | cccagccctt | 7500 |
| ccccaatgac | gtgaccctg  | gggtggctcc | aggtctccag  | tccatgccac  | caaaatctcc | 7560 |
| agattgaggg | tctcccttg  | agtcctgat  | gcctgtccag  | gagctgcccc  | ctgagcaaat | 7620 |
| ctagagtga  | gagggctggg | attgtggcag | taaaagcagc  | cacatttgtc  | tcaggaagga | 7680 |
| aagggaggac | atgagctcca | ggaagggcga | tggcgctctc  | tagtgggcgc  | ctcctgttaa | 7740 |
| tgagcaaaaa | ggggccagga | gagttgagag | atcagggtctg | gccttggact  | aaggctcaga | 7800 |
| tggagaggac | tgaggtgcaa | agagggggct | gaagtagggg  | agtggtcggg  | agagatggga | 7860 |
| ggagcaggta | aggggaagcc | ccagggaggg | cgggggaggg  | tacagcagag  | ctctccactc | 7920 |
| ctcagcattg | acatttgggg | tggctgtgct | agtggggttc  | tgtaaagttgt | aggggtgtta | 7980 |
| gcaccatctg | gggactctac | ccactaaatg | ccagcaggac  | tccctcccca  | agctctaaca | 8040 |
| accaacaatg | tctccagact | ttccaaatgt | cccctggaga  | gcaaaattgc  | ttctggcaga | 8100 |
| atcactgatc | tacgtcagtc | tctaaaagtg | actcatcagc  | gaaatccttc  | acctcttggg | 8160 |
| agaagaatca | caagtgtgag | aggggtagaa | actgcagact  | tcaaaatctt  | tccaaaagag | 8220 |
| ttttacttaa | tcagcagttt | gatgtcccag | gagaagatac  | atttagagtg  | tttagagttg | 8280 |
| atgccacatg | gctgcctgta | cctcacagca | ggagcagagt  | gggttttcca  | agggcctgta | 8340 |
| accacaactg | gaatgacact | cactgggtta | cattacaaag  | tggaatgtgg  | ggaattctgt | 8400 |
| agactttggg | aagggaaatg | tatgacgtga | gcccacagcc  | taaggcagtg  | gacagtccac | 8460 |
| tttgaggctc | tcaccatcta | ggagacatct | cagccatgaa  | catagccaca  | tctgtcatta | 8520 |

|  | 1979-1980 | 1980-1981 | 1981-1982 | 1982-1983 | 1983-1984 | 1984-1985 | 1985-1986 | 1986-1987 | 1987-1988 | 1988-1989 | 1989-1990 | 1990-1991 | 1991-1992 | 1992-1993 | 1993-1994 | 1994-1995 | 1995-1996 | 1996-1997 | 1997-1998 | 1998-1999 | 1999-2000 | 2000-2001 | 2001-2002 | 2002-2003 | 2003-2004 | 2004-2005 | 2005-2006 | 2006-2007 | 2007-2008 | 2008-2009 | 2009-2010 | 2010-2011 | 2011-2012 | 2012-2013 | 2013-2014 | 2014-2015 | 2015-2016 | 2016-2017 | 2017-2018 | 2018-2019 | 2019-2020 | 2020-2021 | 2021-2022 | 2022-2023 | 2023-2024 | 2024-2025 | 2025-2026 | 2026-2027 | 2027-2028 | 2028-2029 | 2029-2030 | 2030-2031 | 2031-2032 | 2032-2033 | 2033-2034 | 2034-2035 | 2035-2036 | 2036-2037 | 2037-2038 | 2038-2039 | 2039-2040 | 2040-2041 | 2041-2042 | 2042-2043 | 2043-2044 | 2044-2045 | 2045-2046 | 2046-2047 | 2047-2048 | 2048-2049 | 2049-2050 | 2050-2051 | 2051-2052 | 2052-2053 | 2053-2054 | 2054-2055 | 2055-2056 | 2056-2057 | 2057-2058 | 2058-2059 | 2059-2060 | 2060-2061 | 2061-2062 | 2062-2063 | 2063-2064 | 2064-2065 | 2065-2066 | 2066-2067 | 2067-2068 | 2068-2069 | 2069-2070 | 2070-2071 | 2071-2072 | 2072-2073 | 2073-2074 | 2074-2075 | 2075-2076 | 2076-2077 | 2077-2078 | 2078-2079 | 2079-2080 | 2080-2081 | 2081-2082 | 2082-2083 | 2083-2084 | 2084-2085 | 2085-2086 | 2086-2087 | 2087-2088 | 2088-2089 | 2089-2090 | 2090-2091 | 2091-2092 | 2092-2093 | 2093-2094 | 2094-2095 | 2095-2096 | 2096-2097 | 2097-2098 | 2098-2099 | 2099-2100 | 2100-2101 | 2101-2102 | 2102-2103 | 2103-2104 | 2104-2105 | 2105-2106 | 2106-2107 | 2107-2108 | 2108-2109 | 2109-2110 | 2110-2111 | 2111-2112 | 2112-2113 | 2113-2114 | 2114-2115 | 2115-2116 | 2116-2117 | 2117-2118 | 2118-2119 | 2119-2120 | 2120-2121 | 2121-2122 | 2122-2123 | 2123-2124 | 2124-2125 | 2125-2126 | 2126-2127 | 2127-2128 | 2128-2129 | 2129-2130 | 2130-2131 | 2131-2132 | 2132-2133 | 2133-2134 | 2134-2135 | 2135-2136 | 2136-2137 | 2137-2138 | 2138-2139 | 2139-2140 | 2140-2141 | 2141-2142 | 2142-2143 | 2143-2144 | 2144-2145 | 2145-2146 | 2146-2147 | 2147-2148 | 2148-2149 | 2149-2150 | 2150-2151 | 2151-2152 | 2152-2153 | 2153-2154 | 2154-2155 | 2155-2156 | 2156-2157 | 2157-2158 | 2158-2159 | 2159-2160 | 2160-2161 | 2161-2162 | 2162-2163 | 2163-2164 | 2164-2165 | 2165-2166 | 2166-2167 | 2167-2168 | 2168-2169 | 2169-2170 | 2170-2171 | 2171-2172 | 2172-2173 | 2173-2174 | 2174-2175 | 2175-2176 | 2176-2177 | 2177-2178 | 2178-2179 | 2179-2180 | 2180-2181 | 2181-2182 | 2182-2183 | 2183-2184 | 2184-2185 | 2185-2186 | 2186-2187 | 2187-2188 | 2188-2189 | 2189-2190 | 2190-2191 | 2191-2192 | 2192-2193 | 2193-2194 | 2194-2195 | 2195-2196 | 2196-2197 | 2197-2198 | 2198-2199 | 2199-2200 | 2200-2201 | 2201-2202 | 2202-2203 | 2203-2204 | 2204-2205 | 2205-2206 | 2206-2207 | 2207-2208 | 2208-2209 | 2209-2210 | 2210-2211 | 2211-2212 | 2212-2213 | 2213-2214 | 2214-2215 | 2215-2216 | 2216-2217 | 2217-2218 | 2218-2219 | 2219-2220 | 2220-2221 | 2221-2222 | 2222-2223 | 2223-2224 | 2224-2225 | 2225-2226 | 2226-2227 | 2227-2228 | 2228-2229 | 2229-2230 | 2230-2231 | 2231-2232 | 2232-2233 | 2233-2234 | 2234-2235 | 2235-2236 | 2236-2237 | 2237-2238 | 2238-2239 | 2239-2240 | 2240-2241 | 2241-2242 | 2242-2243 | 2243-2244 | 2244-2245 | 2245-2246 | 2246-2247 | 2247-2248 | 2248-2249 | 2249-2250 | 2250-2251 |
|--|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
|--|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|

|             |             |             |             |            |            |       |
|-------------|-------------|-------------|-------------|------------|------------|-------|
| gaaaaacatgt | tttattaaga  | ggaaaaaatct | aggctagaag  | tgctttatgc | tcttttttct | 8580  |
| ctttatgttc  | aaattcatat  | acttttagat  | cattcccttaa | agaagaatct | atccccctaa | 8640  |
| gtaaatgtta  | tcactgactg  | gatagtgttg  | gtgtctcact  | cccaaccctt | gtgtggtgac | 8700  |
| agtgccctgc  | ttccccagcc  | ctgggccctc  | tctgattcct  | gagagctttg | ggtgctcctt | 8760  |
| cattaggagg  | aagagaggaa  | gggtgttttt  | aatattctca  | ccattcaccc | atccacctct | 8820  |
| tagacactgg  | gaagaatcag  | ttgccactc   | ttggatttga  | tcctcgaatt | aatgacctct | 8880  |
| at ttctgtcc | cttgtccatt  | tcaacaatgt  | gacaggccta  | agaggtgcct | tctccatgtg | 8940  |
| at ttttgagg | agaaggttct  | caagataagt  | tttctcacac  | ctctttgaat | tacctccacc | 9000  |
| tgtgtcccca  | tcaccattac  | cagcagcatt  | tggacccttt  | ttctgttagt | cagatgcttt | 9060  |
| ccacctcttg  | aggggtgtata | ctgtatgctc  | tctacacagg  | aatatgcaga | ggaaatagaa | 9120  |
| aaaggggaaat | cgcattacta  | ttcagagaga  | agaagacctt  | tatgtgaatg | aatgagagtc | 9180  |
| taaaatccta  | agagagccca  | tataaaatta  | ttaccagtgc  | taaaactaca | aaagttacac | 9240  |
| taacagtaaa  | ctagaataat  | aaaacatgca  | tcacagtgc   | tggtaaagct | aaatcagata | 9300  |
| tttttttctt  | agaaaaagca  | ttccatgtgt  | gttgcaagtga | tgacaggagt | gcccttcagt | 9360  |
| caatatgctg  | cctgtaattt  | ttgttccttg  | gcagaatgta  | ttgtcttttc | tccctttaaa | 9420  |
| tcttaaatgc  | aaaactaaag  | gcagctcctg  | ggccccctcc  | ccaaagtcag | ctgcctgcaa | 9480  |
| ccagccccac  | gaagagcaga  | ggcctgagct  | tccctggcca  | aaataggggg | ctagggagct | 9540  |
| taaccttgct  | cgataaagct  | gtgttcccag  | aatgtcgctc  | ctgttcccag | gggcaccagc | 9600  |
| ctggaggggtg | gtgagcctca  | ctgggtggcct | gatgcttacc  | ttgtgccctc | acaccagtgg | 9660  |
| tcactggaac  | cttgaacact  | tggctgtcgc  | ccggatctgc  | agatgtcaag | aacttctgga | 9720  |
| agtcaaatta  | ctgcccactt  | ctccagggca  | gatacctgtg  | aacatccaaa | accatgccac | 9780  |
| agaaccctgc  | ctgggggtcta | caacacatat  | ggactgtgag  | caccaagtcc | agccctgaat | 9840  |
| ctgtgaccac  | ctgccaagat  | gcccctaact  | gggatccacc  | aatcactgca | catggcaggc | 9900  |
| agcgaggctt  | ggaggtgctt  | cgccacaagg  | cagccccaat  | ttgctgggag | tttcttggca | 9960  |
| cctggtagtg  | gtgaggagcc  | ttgggaccct  | caggattact  | ccccttaagc | atagtgggga | 10020 |
| cccttctgca  | tccccagcag  | gtgccccgct  | cttcagagcc  | tctctctctg | aggtttaccc | 10080 |
| agaccctgc   | accaatgaga  | ccatgctgaa  | gcctcagaga  | gagagatgga | gctttgacca | 10140 |
| ggagccgctc  | ttccttgagg  | gccagggcag  | ggaaagcagg  | aggcagcacc | aggagtggga | 10200 |

# FIGURE 5G

|  |       |
|--|-------|
| acaccagtgt ctaagcccct gatgagaaca ggggtggtctc tcccatatgc ccataccagg | 10260 |
| cctgtgaaca gaatcctcct tctgcagtga caatgtctga gaggacgaca tgtttcccag  | 10320 |
| cctaacgtgc agccatgccc atctaccac tgcctactgc aggacagcac caaccaggga   | 10380 |
| gctgggaagc tgggagaaga catggaatac ccatggcttc tcaccttcct ccagtccagt  | 10440 |
| gggcaccatt tatgcctagg acaccacct gccggcccca ggctcttaag agttaggtca   | 10500 |
| cctaggtgcc tctgggaggc cgaggcagga gaattgcttg aaccgggag gcagaggtg    | 10560 |
| cagtgagccg agatcacacc actgcactcc agcctgggtg acagaatgag actctgtctc  | 10620 |
| aaaaaaaaag agaaagatag catcagtggc taccaagggc taggggcagg ggaaggtgga  | 10680 |
| gagttaatga ttaatagtat gaagtttcta tgtgagatga tgaaaatgtt ctggaaaaaa  | 10740 |
| aaatatagtg gtgaggatgt agaatatgtt gaatataatt aacggcattt aattgtacac  | 10800 |
| ttaacatgat taatgtggca tattttatct tatgtatttg actacatcca agaaacactg  | 10860 |
| ggagagggaa agcccaccat gtaaaataca cccaccctaa tcagatagtc ct cattgtac | 10920 |
| ccaggtagag gccctcatg acctgcacag gaataactaa ggatttaagg acatgaggct   | 10980 |
| tcccagccaa ctgcagggtgc acaacataaa tgtatctgca aacagactga gagtaaagct | 11040 |
| gggggcacaa acctcagcac tgccaggaca cacacccttc tcgtggattc tgactttatc  | 11100 |
| tgacccggcc cactgtccag atcttgttgt gggattggga caaggagggt cataaagcct  | 11160 |
| gtccccaggg cactctgtgt gagcacacga gacctccca cccccacc gttaggtctc     | 11220 |
| cacacataga tctgaccatt aggcattgtg aggaggactc tagcgcgggc tcagggatca  | 11280 |
| caccagagaa tcaggtagag agaggaagac ggggctcgag gagctgatgg atgacacaga  | 11340 |
| gcagggttcc tgcagtccac aggtccagct caccctggtg taggtgcccc atccccctga  | 11400 |
| tccaggcatc cctgacacag ctccctcccg gagcctctc ccagggtgaca catcagggtc  | 11460 |
| cctcactcaa gctgtccaga gagggcagca ccttgacag cgccccccc acttcactct    | 11520 |
| tctccctca cagggtcag ggctcagggc tcaagtctca gaacaaatgg cagaggccag    | 11580 |
| tgagcccaga gatggtgaca gggcaatgat ccaggggcag ctgcctgaaa cgggagcagg  | 11640 |
| tgaagccaca gatgggagaa gatggttcag gaagaaaaat ccaggaatgg gcaggagagg  | 11700 |
| agaggaggac acaggctctg tggggctgca gcccaggatg ggactaagtg tgaagacatc  | 11760 |
| tcagcagggtg aggccaggtc ccatgaacag agaagcagct cccacctccc ctgatgcacg | 11820 |
| gacacacaga gtgtgtggtg ctgtgcccc agagtggggc tctctgttc tgggtccccag   | 11880 |
| ggagtgagaa gtgaggttga cttgtccctg ctctctctctg ctacccaac attcaccttc  | 11940 |

# FIGURE 5H

|  |       |
|--|-------|
| tcctcatgcc cctctctctc aaatatgatt tggatctatg tccccgccca aatctcatgt  | 12000 |
| caaattgtaa accccaatgt tggagggtggg gccttgtgag aagtgattgg ataatgcggg | 12060 |
| tggattttct gctttgatgc tgtttctgtg atagagatct cacatgatct ggttgtttaa  | 12120 |
| aagtgtgtag cacctctccc ctctctctct ctctctctta ctcatgctct gccatgtaag  | 12180 |
| acgttcctgt ttccccctca ccgtccagaa tgattgtaag ttttctgagg cctccccagg  | 12240 |
| agcagaagcc actatgcttc ctgtacaact gcagaatgat gagcgaatta aacctctttt  | 12300 |
| ctttataaat taccagctct cagggtatttc tttatagcaa tgcgaggaca gactaataca | 12360 |
| atcttctact ccagatccc cgcacacgct tagccccaga catcactgcc cctgggagca   | 12420 |
| tgcacagcgc agcctcctgc cgacaaaagc aaagtcacaa aaggtgacaa aaatctgcat  | 12480 |
| ttggggacat ctgattgtga aagaggggagg acagtacact tgtagccaca gagactgggg | 12540 |
| ctcaccgagc tgaaacctgg tagcactttg gcataacatg tgcattgacc gtgttcaatg  | 12600 |
| tctagagatc agtggtgagt aaaacagcct ggtctggggc cgctgctgtc ccacttccc   | 12660 |
| tcctgtccac cagagggcgg cagagttcct cccacctgg agcctccca ggggctgctg    | 12720 |
| acctccctca gccgggcccc cagcccagca gggtcacccc tcaccgggt cacctcgcc    | 12780 |
| cacgtcctcc tcgccctccg agtcctcac acggactctg tcagctctc cctgcagcct    | 12840 |
| atcggccgcc cacctgaggc ttgtcgccg cccacttgag gcctgtcggc tgccctctgc   | 12900 |
| aggcagctcc tgtccctac accccctcct tccccgggt cagctgaaag ggcgtctccc    | 12960 |
| agggcagctc cctgtgatct ccaggacagc tcagtctctc acaggctccg acgcccccta  | 13020 |
| tgctgtcacc tcacagccct gtcattacca ttaactctc agtcccatga agttcactga   | 13080 |
| gcgcctgtct cccggttaca ggaaaactct gtgacaggga ccacgtctgt cctgctctct  | 13140 |
| gtggaatccc agggcccagc ccagtgcctg acacggaaca gatgctccat aaatactggt  | 13200 |
| taaatgtgtg ggagatctct aaaaagaagc atatcacctc cgtgtggccc ccagcagtca  | 13260 |
| gagtctgttc catgtggaca caggggcact ggcaccagca tgggaggagg ccagcaagtg  | 13320 |
| cccgcggctg cccaggaat gaggcctcaa ccccagagc ttcagaaggg aggacagagg    | 13380 |
| cctgcaggga atagatcctc cggcctgacc ctgcagccta atccagagtt cagggtcagc  | 13440 |
| tcacaccagc tcgacctgg tcagcatccc tagggcagtt ccagacaagg ccggaggctct  | 13500 |
| cctcttgccc tccagggggg gacattgcac acagacatca ctcaggaaac ggattcccct  | 13560 |
| ggacaggaac ctggccttgc taaggaagtg gaggtggagc ctggtttcca tcccttgctc  | 13620 |

# FIGURE 5I

|            |            |            |             |            |             |       |
|------------|------------|------------|-------------|------------|-------------|-------|
| caacagaccc | ttctgatctc | tcccacatac | ctgctctgtt  | cctttctggg | tcctatgagg  | 13680 |
| accctgttct | gccaggggtc | cctgtgcaac | tccagactcc  | ctcctggtac | caccatgggg  | 13740 |
| aaggtggggg | gatcacagga | cagtcagcct | cgcagagaca  | gagaccaccc | aggactgtca  | 13800 |
| gggagaacat | ggacaggccc | tgagccgcag | ctcagccaac  | agacacggag | agggaggggtc | 13860 |
| cccctggagc | cttccccaag | gacagcagag | cccagagtca  | cccacctccc | tccaccacag  | 13920 |
| tcctctcttt | ccaggacaca | caagacacct | ccccctccac  | atgcaggatc | tggggactcc  | 13980 |
| tgagacctct | gggcctgggt | ctccatccct | gggtcagtgg  | cggggttggt | ggtactggag  | 14040 |
| acagaggggt | ggtccctccc | cagccaccac | ccagtgagcc  | tttttctagc | ccccagagcc  | 14100 |
| acctctgtca | ccttcctgtt | gggcacatc  | ccaccttccc  | agagccctgg | agagcatggg  | 14160 |
| gagaccggg  | accctgctgg | gtttctctgt | cacaaaggaa  | aataatcccc | ctggtgtgac  | 14220 |
| agacccaagg | acagaacaca | gcagaggtca | gcactgggga  | agacaggttg | tcctcccagg  | 14280 |
| ggatgggggt | ccatccacct | tgccgaaaag | atttgtctga  | ggaactgaaa | atagaaggga  | 14340 |
| aaaaagagga | gggacaaaag | aggcagaaat | gagaggggag  | gggacagagg | acacctgaat  | 14400 |
| aaagaccaca | cccatgaccc | acgtgatgct | gagaagtact  | cctgccctag | gaagagactc  | 14460 |
| agggcagagg | gaggaaggac | agcagaccag | acagtcacag  | cagccttgac | aaaacgttcc  | 14520 |
| tggaactcaa | gctcttctcc | acagaggagg | acagagcaga  | cagcagagac | catggagtct  | 14580 |
| ccctcggccc | ctccccacag | atggtgcac  | ccctggcaga  | ggctcctgct | cacaggtgaa  | 14640 |
| gggaggacaa | cctgggagag | ggtgggagga | gggagctggg  | gtctcctggg | taggacaggg  | 14700 |
| ctgtgagacg | gacagagggc | tcctgttgga | gcctgaatag  | ggaagaggac | atcagagagg  | 14760 |
| gacaggagtc | acaccagaaa | aatcaaattg | aactggaatt  | ggaaaggggc | aggaaaacct  | 14820 |
| caagagttct | attttcctag | ttaattgtca | ctggccacta  | cgttttttaa | aatcataata  | 14880 |
| actgcatcag | atgacacttt | aaataaaaac | ataaccaggg  | catgaaacac | tgtcctcatc  | 14940 |
| cgcctaccgc | ggacattgga | aaataagccc | caggctgtgg  | agggccctgg | gaaccctcat  | 15000 |
| gaactcatcc | acaggaatct | gcagcctgtc | ccaggcaactg | gggtgcaacc | aagatc      | 15056 |

## FIGURE 6A

gaattcagaa ataggggaag gttgaggaag gacactgaac tcaaagggga tacagtgatt 60  
 ggtttatttg tcttctcttc acaacattgg tgctggagga attcccaccc tgaggttatg 120  
 aagatgtctg aacacccaac acatagcact ggagatatga gctcgacaag agtttctcag 180  
 ccacagagat tcacagccta gggcaggagg acactgtacg ccaggcagaa tgacatggga 240  
 attgcgctca cgattggctt gaagaagcaa ggactgtggg aggtgggctt tgtagtaaca 300  
 agagggcagg gtgaactctg attcccatgg gggaatgtga tggtcctgtt acaaattttt 360  
 caagctggca gggaataaaa ccattacgg tgaggacctg tggagggcgg ctgccccaac 420  
 tgataaagga aatagccagg tgggggcctt tccattgta ggggggacat atctggcaat 480  
 agaagccttt gagacccttt aggttacaag tactgaggca gcaaataaaa tgaaatctta 540  
 tttttcaact ttatactgca tgggtgtgaa gatataattg tttctgtaca gggggtgagg 600  
 gaaaggaggg gaggaggaaa gttcctgcag gtctggtttg gtcttgtgat ccaggggggc 660  
 ttggaactat ttaaattaaa ttaaattaaa acaagcgact gttttaaatt aaattaaatt 720  
 aaattaaatt ttactttatt ttatcttaag ttctgggcta catgtgcagg acgtgcagct 780  
 ttgttacata ggtaaactg tgccatggtg gtttgctgta cctatcaacc catcacctag 840  
 gtattaagcc cagcatgcat tagctgtttt tctgacgct ctccctctcc ctgactccca 900  
 caacaggccc cagtgtgtgt tgttcccctc cctgtgtcca tgtgttctca ttgttcagct 960  
 cccacttata agtgagaaca tgtggtgttt ggttttctgt ttctgtgtta gtttgctgag 1020  
 gataatggct tccacctcca tccatgttcc tgcaaaggac gtgatcttat tcttttttat 1080  
 ggttgcatag aaattgtttt tacaaatcca attgatattg tatttaatta caagttaatc 1140  
 taattagcat actagaagag attacagaag atattaggta cattgaatga ggaaatatat 1200  
 aaaataggac gaaggtgaaa tattaggtag gaaaagtata atagttgaaa gaagtaaaaa 1260  
 aaaatatgca tgagtagcag aatgtaaaag aggtgaagaa cgtaatagtg actttttaga 1320  
 ccagattgaa ggacagagac agaaaaattt taaggaattg ctaaaccatg tgagtgttag 1380  
 aagtacagtc aataacatta aagcctcagg aggagaaaag aataggaaaag gaggaaatat 1440  
 gtgaataaat agtagagaca tgtttgatgg attttaaaat atttgaaaga cctcacatca 1500  
 aaggattcat accgtgccat tgaagaggaa gatggaaaag ccaagaagcc agatgaaagt 1560  
 tagaaatatt attggcaaag cttaaattgtt aaaagtccta gagagaaaag atggcagaaa 1620  
 tattggcggg aaagaatgca gaacctagaa tataaattca tccaacagt ttggtagtgt 1680



|  | 1970 | 1971 | 1972 | 1973 | 1974 | 1975 | 1976 | 1977 | 1978 | 1979 | 1980 | 1981 | 1982 | 1983 | 1984 | 1985 | 1986 | 1987 | 1988 | 1989 | 1990 | 1991 | 1992 | 1993 | 1994 | 1995 | 1996 | 1997 | 1998 | 1999 | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 | 2023 | 2024 | 2025 | 2026 | 2027 | 2028 | 2029 | 2030 | 2031 | 2032 | 2033 | 2034 | 2035 | 2036 | 2037 | 2038 | 2039 | 2040 | 2041 | 2042 | 2043 | 2044 | 2045 | 2046 | 2047 | 2048 | 2049 | 2050 | 2051 | 2052 | 2053 | 2054 | 2055 | 2056 | 2057 | 2058 | 2059 | 2060 | 2061 | 2062 | 2063 | 2064 | 2065 | 2066 | 2067 | 2068 | 2069 | 2070 | 2071 | 2072 | 2073 | 2074 | 2075 | 2076 | 2077 | 2078 | 2079 | 2080 | 2081 | 2082 | 2083 | 2084 | 2085 | 2086 | 2087 | 2088 | 2089 | 2090 | 2091 | 2092 | 2093 | 2094 | 2095 | 2096 | 2097 | 2098 | 2099 | 2100 | 2101 | 2102 | 2103 | 2104 | 2105 | 2106 | 2107 | 2108 | 2109 | 2110 | 2111 | 2112 | 2113 | 2114 | 2115 | 2116 | 2117 | 2118 | 2119 | 2120 | 2121 | 2122 | 2123 | 2124 | 2125 | 2126 | 2127 | 2128 | 2129 | 2130 | 2131 | 2132 | 2133 | 2134 | 2135 | 2136 | 2137 | 2138 | 2139 | 2140 | 2141 | 2142 | 2143 | 2144 | 2145 | 2146 | 2147 | 2148 | 2149 | 2150 | 2151 | 2152 | 2153 | 2154 | 2155 | 2156 | 2157 | 2158 | 2159 | 2160 | 2161 | 2162 | 2163 | 2164 | 2165 | 2166 | 2167 | 2168 | 2169 | 2170 | 2171 | 2172 | 2173 | 2174 | 2175 | 2176 | 2177 | 2178 | 2179 | 2180 | 2181 | 2182 | 2183 | 2184 | 2185 | 2186 | 2187 | 2188 | 2189 | 2190 | 2191 | 2192 | 2193 | 2194 | 2195 | 2196 | 2197 | 2198 | 2199 | 2200 | 2201 | 2202 | 2203 | 2204 | 2205 | 2206 | 2207 | 2208 | 2209 | 2210 | 2211 | 2212 | 2213 | 2214 | 2215 | 2216 | 2217 | 2218 | 2219 | 2220 | 2221 | 2222 | 2223 | 2224 | 2225 | 2226 | 2227 | 2228 | 2229 | 2230 | 2231 | 2232 | 2233 | 2234 | 2235 | 2236 | 2237 | 2238 | 2239 | 2240 | 2241 | 2242 | 2243 | 2244 | 2245 | 2246 | 2247 | 2248 | 2249 | 2250 | 2251 | 2252 | 2253 | 2254 | 2255 | 2256 | 2257 | 2258 | 2259 | 2260 | 2261 | 2262 | 2263 | 2264 | 2265 | 2266 | 2267 | 2268 | 2269 | 2270 | 2271 | 2272 | 2273 | 2274 | 2275 | 2276 | 2277 | 2278 | 2279 | 2280 | 2281 | 2282 | 2283 | 2284 | 2285 | 2286 | 2287 | 2288 | 2289 | 2290 | 2291 | 2292 | 2293 | 2294 | 2295 | 2296 | 2297 | 2298 | 2299 | 2300 | 2301 | 2302 | 2303 | 2304 | 2305 | 2306 | 2307 | 2308 | 2309 | 2310 | 2311 | 2312 | 2313 | 2314 | 2315 | 2316 | 2317 | 2318 | 2319 | 2320 | 2321 | 2322 | 2323 | 2324 | 2325 | 2326 | 2327 | 2328 | 2329 | 2330 | 2331 | 2332 | 2333 | 2334 | 2335 | 2336 | 2337 | 2338 | 2339 | 2340 | 2341 | 2342 | 2343 | 2344 | 2345 | 2346 | 2347 | 2348 | 2349 | 2350 | 2351 | 2352 | 2353 | 2354 | 2355 | 2356 | 2357 | 2358 | 2359 | 2360 | 2361 | 2362 | 2363 | 2364 | 2365 | 2366 | 2367 | 2368 | 2369 | 2370 | 2371 | 2372 | 2373 | 2374 | 2375 | 2376 | 2377 | 2378 | 2379 | 2380 | 2381 | 2382 | 2383 | 2384 | 2385 | 2386 | 2387 | 2388 | 2389 | 2390 | 2391 | 2392 | 2393 | 2394 | 2395 | 2396 | 2397 | 2398 | 2399 | 2400 | 2401 | 2402 | 2403 | 2404 | 2405 | 2406 | 2407 | 2408 | 2409 | 2410 | 2411 | 2412 | 2413 | 2414 | 2415 | 2416 | 2417 | 2418 | 2419 | 2420 | 2421 | 2422 | 2 |
|--|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|---|
|--|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|---|

|             |            |            |             |            |            |      |
|-------------|------------|------------|-------------|------------|------------|------|
| gcagctgtag  | ccttttctag | ataatacact | attgtcatac  | atcgcttaag | cgagtgtaaa | 1740 |
| atggtctcct  | cactttattt | atztatatat | ttattttagtt | ttgagatgga | gcctcgctct | 1800 |
| gtctcctagg  | ctggagtgca | atagtgcgat | accactcact  | gcaacctctg | cctcctctgt | 1860 |
| tcaagtgatt  | ttcttacctc | agcctcccga | gtagctggga  | ttacaggtgc | gtgccaccac | 1920 |
| accgggctaa  | tttttgtatt | ttttgtagag | acggggtttt  | gccatgttgg | ccaggctggt | 1980 |
| cttgaactcc  | tgacatcagg | tgatccacct | gccttggcct  | cctaaagtgc | tgggattaca | 2040 |
| ggcatgagcc  | accgtgccc  | accactttat | ttatTTTTTT  | TTTTTatTTT | taaatttcag | 2100 |
| cttctatTTg  | aaatacaggg | ggcacatata | taggattgTT  | acatgggtat | attgaactca | 2160 |
| ggtagtgatc  | atactaccca | acaggtaggt | tttcaaccca  | ctccccctct | tttctcccc  | 2220 |
| attctagtag  | tgtgcagtgT | ctattgttct | catgtttatg  | tctatgtgtg | ctccaggttt | 2280 |
| agctcccacc  | tgtaagtgag | aacgtgtggt | atttgatttt  | ctgtccctgt | gttaattcac | 2340 |
| ttaggattat  | ggcttccagc | tccattcata | ttgctgtaaa  | ggatatgatt | catttttcat | 2400 |
| ggccatgcag  | tattccatat | tgcgtataga | tcacattttc  | tttctTTTTT | TTTTTtgaga | 2460 |
| cggagtcttg  | ctttgctgcc | taggctggag | tgcagtagca  | cgatctcggc | tcactgcaag | 2520 |
| cttcacctcc  | ggggttcacg | tcattcttct | gtctcagctt  | cccaagtagc | tgggactaca | 2580 |
| ggcgcccgcc  | accacgtccg | gctaattttt | ttgtgtgttt  | ttagtagaga | tgggggtttc | 2640 |
| actgtgttag  | ccaggatggt | cttgatctcc | tgacctgtg   | gtccacctgc | ctcggtctcc | 2700 |
| caaagtgctg  | ggattacagg | ggtgagccac | tgcgcccggc  | ccatatatac | cacattttct | 2760 |
| ttaaccaatc  | caccattgat | gggcaactag | gtagattoca  | tggattccac | agttttgcta | 2820 |
| ttgtgtgcag  | tgtggcagta | gacatatgaa | tgaatgtgtc  | TTTTTggtat | aatgatttgc | 2880 |
| attcctttgg  | gtatacagtc | attaatagga | gtgctgggtt  | gaacggtggc | tctgtttaaa | 2940 |
| attcctttgag | aattttccaa | actgtttgcc | atagagagca  | aactaattta | catttccacg | 3000 |
| aacagtatat  | aagcattccc | ttttctccac | agctttgtca  | tcatggtttt | TTTTTTctt  | 3060 |
| tattttaaaa  | aagaatatgt | tgttgTTTTc | ccagggtaca  | tgtgcaggat | gtgcaggttt | 3120 |
| gttacatagg  | tagtaaactg | gagccatggt | ggtttgtctg  | acctgtcaac | ccattacctg | 3180 |
| ggtatgaagc  | cctgcctgca | ttagctcttt | tcctaatagc  | tctcactact | gccccacct  | 3240 |
| cacctgaca   | gggcaaacag | acaacctaca | gaatgggagg  | aaatttttgc | aatctattca | 3300 |
| tctgacaaag  | gtcaagaata | tccagaatct | acaaggaact  | taagcaaatt | tttactTTTT | 3360 |

[illegible]

|             |             |            |             |             |             |      |
|-------------|-------------|------------|-------------|-------------|-------------|------|
| aataatagcc  | actctgactg  | gcgtgaaatg | gtatctcatt  | gtgggttttca | tttgaatttc  | 3420 |
| tctgatgatc  | agtgcacgatg | agcatttttt | catatttggt  | ggctgcttgt  | acgtcttttg  | 3480 |
| agaagtgtct  | cttcatgcct  | tttggccact | ttaatgggat  | tattttttgc  | tttttagttt  | 3540 |
| aagttcctta  | tagattcttg  | atattagact | tcttattgga  | tgcatagttt  | gtgaatactc  | 3600 |
| tcttccattc  | tgtaggttgt  | ctgtttactc | tattgatggc  | ttcttttgct  | gtgccgaagc  | 3660 |
| atcttagttt  | aattagaaac  | cacctgccaa | tttttgtttt  | tgttgcaatt  | gcttttgggg  | 3720 |
| acttagtcat  | aaactctttg  | ccaaggctcg | ggccaagaag  | agtatttcct  | aggttttctt  | 3780 |
| ctagaatttt  | gaaagtctga  | atgtaaacat | ttgcattttt  | aatgcatctt  | gagttagttt  | 3840 |
| ttgtatatgt  | gaaaggctca  | ctctcatttt | ctttccctct  | ttctttcttt  | ctttcttttc  | 3900 |
| tttctttctt  | tctttctttc  | tttctttctt | tctttctttc  | tttctttttg  | tccttctttc  | 3960 |
| tttctttctt  | tctctttctt  | tctctctttc | tttttttttt  | ttgatggagt  | attgctctgt  | 4020 |
| tgcccaggct  | gcagtgcagc  | ggcacgatct | cggtccactg  | caacctctgc  | ctcctggggt  | 4080 |
| caactgattc  | tcctgcatca  | gccttccaag | tagctgggat  | tataggcgcc  | cgccaccacg  | 4140 |
| cccgactaat  | ttttgtattt  | ttagtagaga | cggggttgtg  | ccatgttggc  | caggctgggt  | 4200 |
| tgaaaactcct | gacctcaaac  | gatctgcctg | ccttggcctc  | ccaaagtgct  | gggattacag  | 4260 |
| gtgtgagcca  | ctgtgccag   | ccaagaatgt | cattttctaa  | gagggtccaag | aacctcaaga  | 4320 |
| tattttggga  | ccttgagaag  | agaggaattc | atacaggtat  | tacaagcaca  | gcctaattggc | 4380 |
| aaatctttgg  | catggcttgg  | cttcaagact | ttaggctctt  | aaaagtcgaa  | tccaaaaatt  | 4440 |
| tttataaaag  | ctccagctaa  | gctaccttaa | aaggggcctg  | tatggctgat  | cactcttctt  | 4500 |
| gctatacttt  | acacaaataa  | acaggccaaa | tataatgagg  | ccaaaattta  | ttttgcaaat  | 4560 |
| aaattggtcc  | tgctatgatt  | tactcttggt | aagaacaggg  | aaaatagaga  | aaaatttaga  | 4620 |
| ttgcatctga  | cctttttttc  | tgaattttta | tatgtgccta  | caatttgagc  | taaatcctga  | 4680 |
| attattttct  | ggttgcaaaa  | actctctaaa | gaagaacttg  | gttttcattg  | tcttcgtgac  | 4740 |
| acatttatct  | ggctctttac  | tagaacagct | ttcttgtttt  | tggtgtttcta | gcttgtgtgc  | 4800 |
| cttacagttc  | tactcttcaa  | attattgtta | tgtgtatctc  | atagttttcc  | ttcttttgag  | 4860 |
| aaaactgaag  | ccatgggtatt | ctgaggacta | gagatgactc  | aacagagctg  | gtgaatctcc  | 4920 |
| tcatatgcaa  | tccactgggc  | tcgatctgct | tcaaattgct  | gatgcactgc  | tgctaaagct  | 4980 |
| atacatttaa  | aaccctcact  | aaaggatcag | ggaccatcat  | ggaagaggag  | gaaacatgaa  | 5040 |
| attgtaagag  | ccagattcgg  | ggggtagagt | gtggagggtca | gagcaactcc  | accttgaata  | 5100 |

## FIGURE 6D

agaaggtaaa gcaacctatc ctgaaagcta acctgccatg gtggcttctg attaacctct 5160  
gttctaggaa gactgacagt ttgggtctgt gtcattgccc aaatctcatg ttaaattgta 5220  
atccccagtg ttcgagggtg ggacttggtg gtaggtgatt cggtcatggg agtagatttt 5280  
cttcttttggtg gtgttacagt gatagtgagt gaggttctgt gagatctggt catttaaaag 5340  
tgtgtggccc ctcccctccc tctcttggtc ctccactgc catgtaagat acctgctcct 5400  
gctttgcctt ctaccataag taaaagcccc ctgaggcctc cccagaagca gatgccacca 5460  
tgcttcctgt acagcctgca gaaccatcag ccaattaaac ctcttttctg tataaattac 5520  
cagtcttgag tatctcttta cagcagtgtg agaacggact aatacaaggg tctccaaaat 5580  
tccaagttaa tgtattcttt cttgccaaat agcaggtatt taccataaat cctgtcctta 5640  
ggcacaacaa ccttgatggc atcgtaactc aattgtctta cacattcctt ctgaatgact 5700  
cctcccctat ggcatataag ccctgggtct tgggggataa tggcagaggg gtccaccatc 5760  
ttgtctggct gccacctgag acacggacat ggcttctgtt ggtaagtctc tattaatgt 5820  
ttcttttctaa gaaactggat ttgtcagctt gtttctttgg cctctcagct tcctcagact 5880  
ttggggtagg ttgcacaacc ctgcccacca cgaaacaaat gtttaatatg ataaatatgg 5940  
atagatataa tccacataaa taaaagctct tggagggccc tcaataattg ttaagagtgt 6000  
aaatgtgtcc aaagatggaa aatgtttgag aactactgtc ccagagattt tcctgagttc 6060  
tagagtgtgg gaatatagaa cctggagctt ggcttcttca gcctagaatc aggagtatgg 6120  
ggctgaagtc tgaagcttgg cttcagcagt ttggggttgg cttccggagc acatatttga 6180  
catgttgoga ctgtgatttg gggtttggtg tttgctctga atcctaattg ctgtccttga 6240  
ggcatctaga atctgaaatc tgtggtcaga attctattat cttgagtagg acatctccag 6300  
tcctggttct gccttctagg gctggagtct gtagtcagtg acccggctctg gcatttcaac 6360  
ttcatataca gtgggctatc ttttggtcca tgtttcaacc aaacaaccga ataaaccatt 6420  
agaacctttc cccacttccc tagctgcaat gttaaacctg ggatttctgt ttaataggtt 6480  
catatgaata atttcagcct gatccaactt tacattcctt ctaccgttat tctacacca 6540  
ccttaaaaat gcattcccaa tatattcctt ggattctacc tatatatggg aatcctggct 6600  
ttgccagttt ctagtgcatt aacatactg atttacattc ttttacttta aagtggaaat 6660  
aagagtccct ctgcagagtt caggagttct caagatggcc cttacttctg acatcaattg 6720  
agatttcaag ggagtcgcca agatcatcct caggttcagt gattgctggg agccctcata 6780

# FIGURE 6E

taactcaatg aaagctgtta tgctcatggc tatggtttat tacagcaaaa gaatagagat 6840  
 gaaaatctag caaggggaaga gttgcatggg gcaaagacaa ggagagctcc aagtgcagag 6900  
 attcctgttg ttttctccca gtgggtgcat ggaaagcagt atcttctcca tacaatgatg 6960  
 tgtgataata ttcagtgtat tgccaatcag ggaactcaac tgagccttga ttatattgga 7020  
 gcttggttgc acagacatgt cgaccacctt catggctgaa ctttagtact tagccccctcc 7080  
 agacgtctac agctgatagg ctgtaaccca acattgtcac cataaatcac attgttagac 7140  
 tatccagtgt ggcccaagct cccgtgtaaa cacaggcact ctaaacaggc aggatatttc 7200  
 aaaagcttag agatgacctc ccaggagctg aatgcaaaga cctggcctct ttgggcaagg 7260  
 agaatccttt accgcacact ctccctcaca gggttattgt gaggatcaaa tgtggtcatg 7320  
 tgtgtgagac accagcacat gtctggctgt ggagagtgc ttctatgtgt gctaacattg 7380  
 ctgagtgcata agaaagtatt aggcattggc ttcagcactc acagatgctc atctaatact 7440  
 cacaacatgg ctacaggggtg ggcactacta gcctcatttg acagaggaaa ggactgtgga 7500  
 taagaagggg gtgaccaata ggtcagagtc attctggatg caaggggctc cagaggacca 7560  
 tgattagaca ttgtctgcag agaaattatg gctggatgctc tctgccccgg aaagggggat 7620  
 gcactttcct tgaccccccta tctcagatct tgactttgag gttatctcag acttccctcta 7680  
 tgataaccagg agcccatcat aatctctctg tgcctctctc ccttccctcag tcttactgcc 7740  
 cactcttccc agctccatct ccagctggcc aggtgtagcc acagtaccta actctttgca 7800  
 gagaactata aatgtgtatc ctacagggga gaaaaaaaaa aagaactctg aaagagctga 7860  
 cattttaccg acttgcaaac acataagcta acctgccagt tttgtgctgg tagaactcat 7920  
 gagactcctg ggtcagaggc aaaagatttt attaccaca gctaaggagg cagcatgaac 7980  
 tttgtgttca catttgttca ctttgcccc caattcatat gggatgatca gagcagttca 8040  
 ggtggatgga cacaggggtt tgtggcaaag gtgagcaacc taggcttaga aatcctcaat 8100  
 cttataagaa ggtactagca aacttgtcca gtctttgtat ctgacggaga tattatcttt 8160  
 ataattgggt tgaaagcaga cctactctgg aggaacatat tgtatttatt gtccgaaca 8220  
 gtaaacaaat ctgctgtaaa atagacgtta actttattat ctaaggcagt aagcaaacct 8280  
 agatctgaag gcgataccat cttgcaaggc tatctgctgt acaaatatgc ttgaaaagat 8340  
 ggtccagaaa agaaaacggt attattgcct ttgctcagaa gacacacaga aacataagag 8400  
 aacctggaa aattgtctcc caacactgtt caccagagc cttccactct tgtctgcagg 8460  
 acagtcttaa catcccatca ttagtgtgtc taccacatct ggcttcaccg tgcctaacca 8520

6840 6900 6960 7020 7080 7140 7200 7260 7320 7380 7440 7500 7560 7620 7680 7740 7800 7860 7920 7980 8040 8100 8160 8220 8280 8340 8400 8460 8520

## FIGURE 6F

agatttctag gtccagttcc ccaccatggt tggcagtgcc cactgcca cccagaata 8580  
agggagtgtcagaattccg aggggacatg ggtggggatc agaacttctg ggcttgagtg 8640  
cagagggggc ccatactcct tggttccgaa ggaggaagag gctggaggtg aatgtccttg 8700  
gaggggagga atgtgggttc tgaactctta aatccccaag ggaggagact ggtaagggtcc 8760  
cagcttccga ggtactgacg tgggaatggc ctgagaggtc taagaatccc gtatcctcgg 8820  
gaaggagggg ctgaaattgt gaggggttga gttgcagggg tttgttagct tgagactcct 8880  
tggtgggtcc ctgggaagca aggactggaa ccattggctc cagggtttg tgtgaaggta 8940  
atgggatctc ctgattctca aagggtcaga ggactgagag ttgccatgc tttgatcttt 9000  
ccatctactc cttactccac ttgagggtaa tcacctactc ttctagttcc acaagagtgc 9060  
gcctgcgcga gtataatctg cacatgtgcc atgtcccgag gcctggggca tcatccactc 9120  
atcattcagc atctgcgcta tgcgggcgag gccggcgcca tgacgtcatg tagctgcgac 9180  
tatccctgca gcgcgcctct cccgtcacgt cccaacctg gagctgtgga cgtgcgtccc 9240  
ctggtggatg tggcctgcgt ggtgccaggc cggggccttg tgtccgataa agatcctaga 9300  
accacaggaa accaggactg aaagggtgta gagaatggcc atatgtcgt gtccatgaaa 9360  
tctcaaggac ttctgggttg agggcacagg agcctgaact tacgggttg cccagttcca 9420  
ctgtcctccc aagtgagtct ccagatacg aggactgtg ccagcatcag cttcatctgt 9480  
accacatctt gtaacaggga ctaccaggga ccctgatgaa caccatggtg tgtgcaggaa 9540  
gagggggtga aggcatggac tcctgtgttg tcagagccca gagggggcca tgacgggttg 9600  
ggaggaggct gtggactggc tcgagaagtg ggatgtggtt gtgtttgatt tcctttggcc 9660  
agataaagtg ctggatatag cattgaaaac ggagtatgaa gaccagttag aatggagggt 9720  
caggttgag ttgagttaca gatggggtaa aattctgctt cggatgagtt tggggatttg 9780  
caatctaaag gtggtttggg atggcatggc tttgggatgg aaataggttt gtttttatgt 9840  
tggtctggga ggtgtgggg attgaattgg ggatgaagta ggttagttt tggagataga 9900  
atacatggag ctggctattg catgcgagga tgtgcattag tttggtttga tctttaaata 9960  
aaggaggcta ttagggttgt cttgaattag attaagttgt gttgggttga tgggttgggc 10020  
ttgtgggtga tgtggttga tgggctgtg ttaaattggt ttgggtcagg ttttggttga 10080  
ggttatcatg gggatgagga tatgcttggg acatggattc aggtggttct cattcaagct 10140  
gaggcaaatt tcctttcaga cggtcattcc agggaacgag tggttgtgtg ggggaaatca 10200

# FIGURE 6G

ggccactggc tgtgaatatc cctctatcct ggtcttgaat tgtgattatc tatgtccatt 10260  
 ctgtctcctt cactgtactt ggaattgacg tggctattca gctggaaatg ggggaagatt 10320  
 ttgtcaaatt cttgagacac agctgggtct ggatcagcgt aagccttcct tctgggttta 10380  
 ttgaacagat gaaatcacat tttttttttc aaaatcacag aaatcttata gagttaacag 10440  
 tggactctta taataagagt taacaccagg actcttatto ttgattcttt tctgagacac 10500  
 caaatgaga tttctcaatg ccaccctaata tttttttttt tttttttttt tttttgagac 10560  
 acagtctggg tcttttgctc tgtcactcag gctggagcgc agtgggtgtga tcatagctca 10620  
 ctgaaccctt gacctcctgg acttaaggga tcctcctgct tcagcctcct gagtagatgg 10680  
 ggctacaggt gcttgccacc acacctgggt aattaaattt tttttttttt tttgtagaga 10740  
 aagggtctca ctttggtgcc ctggctgacg ttgaacttct gacttcaagt gattcttcag 10800  
 ccttggtctc ccaaagcact gggattgctg gcatgagcca ctcaccgtgc ctggccttgc 10860  
 gcttaatctt ggagtgtata aacctggctc ctgatatgta gacatttcag tgagaaggag 10920  
 gcattggatt ttgcatgagg acaattctga cctaggaggg cagggtcaaca ggaatccccg 10980  
 ctgtacctgt acgttggtaca ggcatggaga atgaggagtg aggaggccgt accggaaccc 11040  
 catattgttt agtggacatt ggattttgaa ataataggga acttgggtctg ggagagtcac 11100  
 atttctggat tggacaatat gtggtatcac aaggttttat gatgagggag aaatgtatgt 11160  
 ggggaaccat tttctgagtg tggaagtgcg agaatcagag agtagctgaa tgccaacgct 11220  
 tctatttcag gaacatggta agttggaggt ccagctctcg ggctcagacg ggtataggga 11280  
 ccaggaagtc tcacaatccg atcattctga tatttcaggg catattaggt ttgggggtgc 11340  
 aaggaagtac ttgggactta ggcacatgag actttgtatt gaaaatcaat gattggggct 11400  
 ggccgtgggt ctcacgctg taatctcatc actttgggag accgaagtgg gaggatggct 11460  
 tgatctcaag agttggacac cagcctaggc aacatggcca gaccctctct ctacaaaaaa 11520  
 attaaaaatt agctggatgt ggtggtgcat gcttggtgct tcagctatcc tggaggctga 11580  
 gacaggagaa tcggttgagt ctgggagttc aaggctacag ggagctgcga tcacgccgct 11640  
 gcaactccagc ctgggaaaca gactgagact gtctcagaat ttttttaaaa aagaatcagt 11700  
 gatcatccca acccctgttg ctgttcatcc tgagcctgcc ttctctggct ttgttcctta 11760  
 gatcacatct ccatgatcca taggcctgc ccaatctgac ctcacaccgt gggaatgcct 11820  
 ccagactgat ctagtatgtg tggaacagca agtgcctggc ctccctcccc ttccacagct 11880  
 ctgggtgtgg gagggggttg tccagcctcc agcagcatgg ggagggcctt ggtcagcatc 11940

## FIGURE 6H

taggtgcaa cagggcaagg gcggggctct ggagaatgaa ggctttatag ggctcctcag 12000  
ggaggcccc cagccccaaa ctgcaccacc tggcctgga caccggt 12047

bioRxiv preprint doi: <https://doi.org/10.1101/2021.03.10.434444>; this version posted March 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

## FIGURE 7

cgagcggccc ctcagcttcg gcgcccagcc ccgcaaggct cccggtgacc actagagggc 60  
 gggaggagct cctggccagt ggtggagagt ggcaaggaag gaccctaggg ttcacggag 120  
 cccaggttta ctcccttaag tggaaatttc ttccccact cctccttggc tttctccaag 180  
 gaggggaaccc aggtctgtgg aaagtccggc tggggcgggg actgtgggtt caggggagaa 240  
 cggggtgtgg aacgggacag ggagcggtta gaagggtggg gctattccgg gaagtgggtg 300  
 ggggagggag cccaaaacta gcacctagtc cactcattat ccagccctct tatttctcgg 360  
 ccgctctgct tcagtggacc cggggagggc ggggaagtgg agtgggagac ctaggggtg 420  
 gcttcccgac cttgctgtac aggacctcga cctagctggc ttgtttccc atccccacgt 480  
 tagttgttgc cctgaggcta aaactagagc ccaggggccc caagttccag actgcccctc 540  
 cccctcccc cggagccagg gagtgggttg tgaaaggggg aggccagctg gagaacaaac 600  
 gggtagtcag ggggttgagc gattagagcc cttgtaccct acccaggaat ggttggggag 660  
 gaggaggaag aggtaggagg taggggaggg ggcggggttt tgtcacctgt cacctgctcg 720  
 ctgtgcctag ggcgggcggg cggggagtgg ggggaccggt ataaagcggg aggcgcctgt 780  
 gcccgtcca cctctcaagc agccagcgcc tgctgaatc tgttctgccc cctccccacc 840  
 catttcacca ccacatg 858



| Variable                  | Mean | SD   | Min  | Max   | Skewness | Kurtosis | Normality |
|---------------------------|------|------|------|-------|----------|----------|-----------|
| Age                       | 35.2 | 12.5 | 18   | 65    | -0.1     | 3.2      | 0.98      |
| Gender                    | 0.5  | 0.5  | 0    | 1     | 0.0      | 0.0      | 0.99      |
| Education                 | 12.5 | 2.5  | 8    | 16    | -0.2     | 3.5      | 0.97      |
| Income                    | 4500 | 1500 | 1000 | 10000 | 0.5      | 4.0      | 0.95      |
| Health                    | 2.5  | 1.0  | 1    | 4     | -0.3     | 3.8      | 0.96      |
| Stress                    | 3.0  | 1.5  | 1    | 5     | 0.2      | 3.3      | 0.98      |
| Quality of Life           | 4.0  | 1.0  | 2    | 5     | -0.1     | 3.1      | 0.99      |
| Depression                | 1.5  | 1.0  | 0    | 3     | 0.1      | 3.4      | 0.97      |
| Life Satisfaction         | 3.5  | 1.0  | 1    | 5     | -0.2     | 3.6      | 0.96      |
| Resilience                | 2.0  | 1.0  | 0    | 4     | 0.0      | 3.0      | 0.98      |
| Optimism                  | 3.0  | 1.0  | 1    | 5     | -0.1     | 3.2      | 0.99      |
| Gratitude                 | 2.5  | 1.0  | 0    | 4     | 0.1      | 3.3      | 0.97      |
| Forgiveness               | 2.0  | 1.0  | 0    | 4     | 0.0      | 3.0      | 0.98      |
| Compassion                | 2.5  | 1.0  | 0    | 4     | 0.1      | 3.3      | 0.97      |
| Kindness                  | 2.0  | 1.0  | 0    | 4     | 0.0      | 3.0      | 0.98      |
| Generosity                | 2.5  | 1.0  | 0    | 4     | 0.1      | 3.3      | 0.97      |
| Patience                  | 2.0  | 1.0  | 0    | 4     | 0.0      | 3.0      | 0.98      |
| Self-control              | 2.5  | 1.0  | 0    | 4     | 0.1      | 3.3      | 0.97      |
| Emotional Stability       | 3.0  | 1.0  | 1    | 5     | -0.1     | 3.2      | 0.99      |
| Psychological Well-being  | 3.5  | 1.0  | 1    | 5     | -0.2     | 3.6      | 0.96      |
| Life Purpose              | 3.0  | 1.0  | 1    | 5     | -0.1     | 3.2      | 0.99      |
| Meaning in Life           | 3.5  | 1.0  | 1    | 5     | -0.2     | 3.6      | 0.96      |
| Existential Well-being    | 3.0  | 1.0  | 1    | 5     | -0.1     | 3.2      | 0.99      |
| Transcendental Well-being | 3.5  | 1.0  | 1    | 5     | -0.2     | 3.6      | 0.96      |
| Humanistic Well-being     | 3.0  | 1.0  | 1    | 5     | -0.1     | 3.2      | 0.99      |
| Existential Well-being    | 3.5  | 1.0  | 1    | 5     | -0.2     | 3.6      | 0.96      |
| Transcendental Well-being | 3.0  | 1.0  | 1    | 5     | -0.1     | 3.2      | 0.99      |
| Humanistic Well-being     | 3.5  | 1.0  | 1    | 5     | -0.2     | 3.6      | 0.96      |
| Existential Well-being    | 3.0  | 1.0  | 1    | 5     | -0.1     | 3.2      | 0.99      |
| Transcendental Well-being | 3.5  | 1.0  | 1    | 5     | -0.2     | 3.6      | 0.96      |
| Humanistic Well-being     | 3.0  | 1.0  | 1    | 5     | -0.1     | 3.2      | 0.99      |
| Existential Well-being    | 3.5  | 1.0  | 1    | 5     | -0.2     | 3.6      | 0.96      |
| Transcendental Well-being | 3.0  | 1.0  | 1    | 5     | -0.1     | 3.2      | 0.99      |
| Humanistic Well-being     | 3.5  | 1.0  | 1    | 5     | -0.2     | 3.6      | 0.96      |
| Existential Well-being    | 3.0  | 1.0  | 1    | 5     | -0.1     | 3.2      | 0.99      |
| Transcendental Well-being | 3.5  | 1.0  | 1    | 5     | -0.2     | 3.6      | 0.96      |
| Humanistic Well-being     | 3.0  | 1.0  | 1    | 5     | -0.1     | 3.2      | 0.99      |
| Existential Well-being    | 3.5  | 1.0  | 1    | 5     | -0.2     | 3.6      | 0.96      |
| Transcendental Well-being | 3.0  | 1.0  | 1    | 5     | -0.1     | 3.2      | 0.99      |
| Humanistic Well-being     | 3.5  | 1.0  | 1    | 5     | -0.2     | 3.6      | 0.96      |
| Existential Well-being    | 3.0  | 1.0  | 1    | 5     | -0.1     | 3.2      | 0.99      |
| Transcendental Well-being | 3.5  | 1.0  | 1    | 5     | -0.2     | 3.6      | 0.96      |
| Humanistic Well-being     | 3.0  | 1.0  | 1    | 5     | -0.1     | 3.2      | 0.99      |
| Existential Well-being    | 3.5  | 1.0  | 1    | 5     | -0.2     | 3.6      | 0.96      |
| Transcendental Well-being | 3.0  | 1.0  | 1    | 5     | -0.1     | 3.2      | 0.99      |
| Humanistic Well-being     | 3.5  | 1.0  | 1    | 5     | -0.2     | 3.6      | 0.96      |
| Existential Well-being    | 3.0  | 1.0  | 1    | 5     | -0.1     | 3.2      | 0.99      |
| Transcendental Well-being | 3.5  | 1.0  | 1    | 5     | -0.2     | 3.6      | 0.96      |
| Humanistic Well-being     | 3.0  | 1.0  | 1    | 5     | -0.1     | 3.2      | 0.99      |
| Existential Well-being    | 3.5  | 1.0  | 1    | 5     | -0.2     | 3.6      | 0.96      |
| Transcendental Well-being | 3.0  | 1.0  | 1    | 5     | -0.1     | 3.2      | 0.99      |
| Humanistic Well-being     | 3.5  | 1.0  | 1    | 5     | -0.2     | 3.6      | 0.96      |
| Existential Well-being    |      |      |      |       |          |          |           |

| Variable                    | Mean | SD   | Min  | Max   | Skewness | Kurtosis | Normality |
|-----------------------------|------|------|------|-------|----------|----------|-----------|
| Age                         | 35.2 | 12.5 | 18   | 65    | -0.1     | 3.2      | 0.98      |
| Gender                      | 0.5  | 0.5  | 0    | 1     | 0.0      | 0.0      | 0.99      |
| Education                   | 12.5 | 2.5  | 8    | 16    | -0.2     | 3.5      | 0.97      |
| Income                      | 4500 | 1500 | 1000 | 10000 | 0.5      | 4.0      | 0.95      |
| Health                      | 2.5  | 1.0  | 1    | 4     | -0.3     | 3.8      | 0.96      |
| Stress                      | 3.0  | 1.5  | 1    | 5     | 0.2      | 3.3      | 0.98      |
| Quality of Life             | 4.0  | 1.0  | 1    | 5     | -0.1     | 3.2      | 0.99      |
| Depression                  | 2.0  | 1.0  | 1    | 4     | 0.1      | 3.4      | 0.97      |
| Life Satisfaction           | 3.5  | 1.0  | 1    | 5     | -0.2     | 3.5      | 0.98      |
| Physical Activity           | 1.5  | 1.0  | 0    | 3     | 0.3      | 3.6      | 0.96      |
| Social Support              | 2.5  | 1.0  | 1    | 4     | -0.1     | 3.2      | 0.99      |
| Work Satisfaction           | 3.0  | 1.0  | 1    | 4     | 0.0      | 3.3      | 0.98      |
| Life Expectancy             | 75.0 | 5.0  | 60   | 90    | -0.1     | 3.2      | 0.99      |
| Healthcare Access           | 4.0  | 1.0  | 1    | 5     | -0.2     | 3.5      | 0.97      |
| Environmental Quality       | 3.0  | 1.0  | 1    | 4     | 0.1      | 3.4      | 0.98      |
| Community Engagement        | 2.0  | 1.0  | 1    | 4     | 0.2      | 3.6      | 0.96      |
| Urbanization                | 0.5  | 0.5  | 0    | 1     | 0.0      | 0.0      | 0.99      |
| Population Density          | 1000 | 500  | 0    | 2000  | 0.5      | 4.0      | 0.95      |
| Infrastructure Development  | 3.0  | 1.0  | 1    | 4     | -0.1     | 3.2      | 0.99      |
| Economic Growth             | 5.0  | 1.0  | 1    | 6     | 0.0      | 3.3      | 0.98      |
| Social Equality             | 4.0  | 1.0  | 1    | 5     | -0.2     | 3.5      | 0.97      |
| Political Stability         | 3.5  | 1.0  | 1    | 4     | 0.1      | 3.4      | 0.98      |
| Environmental Awareness     | 2.5  | 1.0  | 1    | 4     | 0.2      | 3.6      | 0.96      |
| Healthcare Expenditure      | 1.0  | 0.5  | 0    | 2     | 0.3      | 3.7      | 0.95      |
| Urban Planning              | 3.0  | 1.0  | 1    | 4     | -0.1     | 3.2      | 0.99      |
| Community Health            | 2.0  | 1.0  | 1    | 4     | 0.1      | 3.4      | 0.97      |
| Life Quality Index          | 4.5  | 1.0  | 1    | 5     | -0.2     | 3.5      | 0.98      |
| Healthcare Access Index     | 3.0  | 1.0  | 1    | 4     | 0.0      | 3.3      | 0.98      |
| Environmental Quality Index | 2.5  | 1.0  | 1    | 4     | 0.1      | 3.4      | 0.97      |
| Community Engagement Index  | 2.0  | 1.0  | 1    | 4     | 0.2      | 3.6      | 0.96      |
| Work Satisfaction Index     | 3.0  | 1.0  | 1    | 4     | 0.0      | 3.3      | 0.98      |
| Life Expectancy Index       | 75.0 | 5.0  | 60   | 90    | -0.1     | 3.2      | 0.99      |
| Healthcare Access Index     | 4.0  | 1.0  | 1    | 5     | -0.2     | 3.5      | 0.97      |
| Environmental Quality Index | 3.0  | 1.0  | 1    | 4     | 0.1      | 3.4      | 0.98      |
| Community Engagement Index  | 2.5  | 1.0  | 1    | 4     | 0.2      | 3.6      | 0.96      |
| Work Satisfaction Index     | 3.0  | 1.0  | 1    | 4     | 0.0      | 3.3      | 0.98      |
| Life Expectancy Index       | 75.0 | 5.0  | 60   | 90    | -0.1     | 3.2      | 0.99      |
| Healthcare Access Index     | 4.0  | 1.0  | 1    | 5     | -0.2     | 3.5      | 0.97      |
| Environmental Quality Index | 3.0  | 1.0  | 1    | 4     | 0.1      | 3.4      | 0.98      |
| Community Engagement Index  | 2.5  | 1.0  | 1    | 4     | 0.2      | 3.6      | 0.96      |
| Work Satisfaction Index     | 3.0  | 1.0  | 1    | 4     | 0.0      | 3.3      | 0.98      |
| Life Expectancy Index       | 75.0 | 5.0  | 60   | 90    | -0.1     | 3.2      | 0.99      |
| Healthcare Access Index     | 4.0  | 1.0  | 1    | 5     | -0.2     | 3.5      | 0.97      |
| Environmental Quality Index | 3.0  | 1.0  | 1    | 4     | 0.1      | 3.4      | 0.98      |
| Community Engagement Index  | 2.5  | 1.0  | 1    | 4     | 0.2      | 3.6      | 0.96      |
| Work Satisfaction Index     | 3.0  | 1.0  | 1    | 4     | 0.0      | 3.3      | 0.98      |
| Life Expectancy Index       | 75.0 | 5.0  | 60   | 90    | -0.1     | 3.2      | 0.99      |
| Healthcare Access Index     | 4.0  | 1.0  | 1    | 5     | -0.2     | 3.5      | 0.97      |
| Environmental Quality Index | 3.0  | 1.0  | 1    | 4     | 0.1      | 3.4      | 0.98      |
| Community Engagement Index  | 2.5  | 1.0  | 1    | 4     | 0.2      | 3.6      | 0.96      |
| Work Satisfaction Index     | 3.0  | 1.0  | 1    | 4     | 0.0      | 3.3      | 0.98      |
| Life Expectancy Index       | 75.0 | 5.0  | 60   | 90    | -0.1     | 3.2      | 0.99      |
| Healthcare Access Index     | 4.0  | 1.0  | 1    |       |          |          |           |

## FIGURE 9

```

g atg acc ggc tca acc atc gcg ccc aca acg gac tat cgc aac acc act 49
Met Thr Gly Ser Thr Ile Ala Pro Thr Thr Asp Tyr Arg Asn Thr Thr
  1           5           10           15

gct acc gga cta aca tct gcc cta aat tta ccc caa gtt cat gcc ttt 97
Ala Thr Gly Leu Thr Ser Ala Leu Asn Leu Pro Gln Val His Ala Phe
      20           25           30

gtc aat gac tgg gcg agc ttg gac atg tgg tgg ttt tcc ata gcg ctt 145
Val Asn Asp Trp Ala Ser Leu Asp Met Trp Trp Phe Ser Ile Ala Leu
      35           40           45

atg ttt gtt tgc ctt att att atg tgg ctt att tgt tgc cta aag cgc 193
Met Phe Val Cys Leu Ile Ile Met Trp Leu Ile Cys Cys Leu Lys Arg
      50           55           60

aga cgc gcc aga ccc ccc atc tat agg cct atc att gtg ctc aac cca 241
Arg Arg Ala Arg Pro Pro Ile Tyr Arg Pro Ile Ile Val Leu Asn Pro
      65           70           75

cac aat gaa aaa att cat aga ttg gac ggt ctg aaa cca tgt tct ctt 289
His Asn Glu Lys Ile His Arg Leu Asp Gly Leu Lys Pro Cys Ser Leu
      85           90           95

ctt tta cag tat gat taa
Leu Leu Gln Tyr Asp
      100
307

```